

Investigating the large scale structure of the universe: The cosmic web as a probe for cosmological parameters.

Håkon Tansem



Thesis submitted for the degree of
Master of Science in Astronomy

Institute of Theoretical Astrophysics
University of Oslo

31st May 2021

Copyright © 2021, Håkon Tansem

This work, entitled “Investigating the large scale structure of the universe: The cosmic web as a probe for cosmological parameters.” is distributed under the terms of the Public Library of Science Open Access License, a copy of which can be found at <http://www.publiclibraryofscience.org>.

Abstract

The large scale structure of the universe provides independent sets of experiments, in addition to established experiments such as CMB analysis and type Ia supernova, for extracting cosmological information. Due to redshift space distortions, anisotropy is introduced when observing the matter distribution in the universe. This allows for the measurement of cosmological parameters such as the energy density contributions and growth rate of structure. Future cosmological surveys, such as Euclid, is expected to greatly increase the number of galaxies having their redshifts measured thereby increasing the size of datasets. In preparation for the analysis of future observational datasets, the methodology has to be thoroughly tested in order to extract as much cosmological information from the density field as possible.

In this thesis, I derive and implement a model for the radial velocity of matter around cosmological filaments using linear perturbation theory. A proposed improvement to the radial velocity of matter surrounding voids is also implemented and tested. This model, together with the regular radial velocity derived from linear theory, is implemented into the void-galaxy cross-correlation function in redshift space using a linear bias approximation for the overdensity profile. In addition, the regular radial velocity proposed by linear theory is also implemented into the void-galaxy cross-correlation function in redshift space using an arbitrary dark matter overdensity profile. Using the MultiDark simulation suite, these models are compared with data from numerical calculations and their performance is assessed.

I find the model to provide promising results when predicting the radial velocity of matter around cosmological filaments. Although the amplitude of the velocity predicted by linear theory is not entirely accurate, it manages to replicate the shape of the radial velocity derived from the dataset. More work is needed to see if this can prove to be a full-fledged addition to LSS studies. When applying fits to the void-galaxy correlation function in redshift space, I find the regular velocity model derived from linear theory to perform consistently better than the model with the proposed improvement when fitting the Alcock-Paczyński parameter ϵ . Applying cuts to the dataset, including only voids with an effective radius $r > 40\text{Mpc}/h$, I find the regular model to also perform better when fitting the growth rate f/b . Lastly, I find the parameter fits to the model using the dark matter overdensity profile to consistently provide accurate fits for ϵ . However, the model does not provide accurate fits for the normalization of the dark matter overdensity profile even with cuts applied.

Acknowledgments

First of all, I would like to thank my supervisor Hans Arnold Winther for providing me with this fun and interesting project. Thank you for providing great help throughout a year that has been strange for everyone and for taking the time to answer all of my questions. I have learned a lot! I would also like to thank all of my fellow students here in the basement at ITA for providing a good study environment with many interesting discussions. Some related to astronomy and some non-related. A special thanks to Nils-Ole, with whom I have known and worked with since the first semester of my bachelor's degree. I would also like to thank my family for supporting me throughout all of my studies here at the university. Lastly, i want to thank my childhood friends at home for making me think about other things than physics and astronomy once in a while.

Contents

Abstract	iii
Acknowledgments	v
List of Figures	ix
1 Introduction	1
2 Background Theory	5
2.1 History of cosmology and how we model the universe	5
2.1.1 Distances in the universe and cosmological redshift	5
2.1.2 Modelling the evolution of the universe and the Λ CDM model. . .	7
2.1.3 Einstein-de Sitter model	11
2.1.4 Growth of matter perturbations in linear perturbation theory. . .	12
2.1.5 Spherical collapse models	14
2.1.6 Two point correlation functions and the matter power spectrum.	15
2.1.7 The cosmic web	17
2.1.8 Real- vs Redshift space	19
2.2 Cosmological N-body simulations	21
2.2.1 Pairwise summation	21
2.2.2 Particle mesh	23
2.2.3 Tree codes and Barnes-Hut simulations	24
2.2.4 Initial conditions	26
2.3 Statistics	26
2.3.1 Bayes theorem	26
2.3.2 Parameter estimation	27
2.3.3 Sampling and Metropolis-Hastings MCMC	28
3 Codes utilized - Identifying voids and filaments from a particle distribution	31
3.1 DisPerSE-"Discrete Persistent Structures Extractor"	31
3.1.1 Morse Theory	31
3.1.2 Discrete Morse theory	32
3.1.3 Delaunay Tesselation Field Estimator	34

3.1.4	Topological Persistence	34
3.2	REVOLVER-REal-space VOId Locations from surVEy Reconstruction	35
3.2.1	ZOBOV-ZONes BOrdering on Voidness	37
3.2.2	Redshift space reconstruction	37
4	Dataset - Multidark Simulation	41
4.1	Halo finding using Rockstar	41
4.2	Applying mass cuts to the dataset.	43
5	Method	45
5.1	Radial velocity components in linear theory	45
5.1.1	Radial velocity component for filaments	45
5.1.2	Radial velocity component for voids	46
5.1.3	Minor correction term to velocity around voids.	47
5.2	Numerical measurement of correlation functions and multipole expansions.	48
5.3	Void Analysis	49
5.3.1	Density profile	49
5.3.2	Velocity profile and velocity dispersion	50
5.3.3	Void-galaxy cross correlation function.	51
5.3.4	Expanding the model to account for velocity dispersion	54
5.4	Filament Analysis	55
5.4.1	Masking the simulation volume	55
5.4.2	Density profile	55
5.4.3	Velocity profile	56
5.5	Parametrizing distance scales.	57
5.6	Calibrating dark matter profiles.	58
5.7	Likelihood analysis.	60
5.7.1	Estimating the covariance matrix	60
5.7.2	Fitting parameters to data.	61
6	Results/Discussion	63
6.1	Filament analysis	63
6.1.1	Effect of choosing persistence ratio on dataset	63
6.1.2	Density profiles	67
6.1.3	Radial velocity profiles	72
6.2	Void analysis	77
6.2.1	Histograms	77
6.2.2	Density profiles	80
6.2.3	Radial velocity profiles	89
6.3	Parameter fits	96
6.3.1	Linear bias approximation	97
6.3.2	Linear bias approximation with correction term	105
6.3.3	Dark matter profile fits	114
6.3.4	Correlation function void-galaxy ξ_{vg}	121

7	Concluding remarks	129
7.1	Summary of main results	129
7.2	Suggestions for future work	131
	Bibliography	133

List of Figures

2.1	The SDSS (Sloan Digital Sky survey) map of the universe. Each dot represents a galaxy and the color scale represents the local density. Downloaded from the SDSS website https://www.sdss.org/science/orangepie/ . Image Credit: M. Blanton and SDSS.	8
2.2	A simple illustration of the spherical top hat model.	14
2.3	The matter power spectrum inferred from different cosmological surveys. Image credit: "ESA and the Planck Collaboration" [1].	16
2.4	Figure showing snapshots of the Millennium simulation with 10^{10} particles for two different redshifts. Figures a) and b) is taken at $z = 0$ while figures c) and d) is taken at $z = 18.3$. Image credit: Downloaded from the website of the Millennium Simulation[2] (https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/).	18
2.5	Figure illustrating a collapsing halo in realspace and redshiftspace. The vector pointers represent the direction of the peculiar velocity. In this illustration both circles are meant to contain the same number density of galaxies thereby giving the apparent overdensity appear much higher in redshiftspace than in realspace. The observer sitting at the bottom of the page with the line of sight direction being parallel to the plane spanned by the circle and the ellipse.	20
2.6	Illustration of a Barnes-Hut tree for 100 particles in 2D space. Image downloaded from https://commons.wikimedia.org/wiki/File:2D_Quad-Tree_partitioning_of_100_bodies.png . The file is licenced under the following license https://creativecommons.org/licenses/by-sa/3.0/deed.en	25
3.1	Example on how DisPerSE uses persistence pairs to simplify topological features. On the left is the unmodified function where the two leftmost critical points forms a persistence pair. The value of the minima is increased while the value of the maxima is increased. This is done until the value of the minima surpasses that of the maximum and thus the persistence pair is cancelled. This effectively smoothes out features of the function.	36

5.1	Illustration of a cylindrical filament with longitudinal axis L and radial velocity component v_r	45
5.2	Illustration of a spherical void with radial velocity pointing outwards from the center of the void.	47
5.3	Figure illustrating the system used in this analysis. This figure is drawn in real space thereby all vectors denoted as \mathbf{r} . Transformation to redshift space will distort the halo positions and the corresponding system will have vectors denoted as \mathbf{s}	49
5.4	Figure illustrating visually how the halo catalogue is masked. On the left is a scatter plot of the halo catalogue without a mask. The right figure illustrates the masking of the catalogue where the two red dots are the start- and end-points of the filament segment. Here one can see that only particles bound by the spine of the filament segment itself are kept. The dataset used for this illustration is the MD1 dataset with $\sigma = 1$ given as persistence threshold to DisPerSE. All units on the axes are in Mpc/h.	56
5.5	Marginalized constraints on Ω_m and Ω_Λ inferred from different cosmological surveys. This figure illustrates that the study of voids is a good addition to other experiments in contributing to providing constraints for these cosmological parameters. Figure is copied from [3].	59
6.1	Figure showing slices of the MD1 dataset in the xy -plane for $950\text{Mpc/h} \leq z \leq 1050\text{Mpc/h}$. Here the persistence threshold given to DisPerSE is varied between no cuts, $\sigma = 1$, $\sigma = 2$ and $\sigma = 3$. Each blue dot represents a halo particle while all lines represents filaments assigned by DisPerSE.	64
6.2	Figure showing slices of the MD2 dataset in the xy -plane for $990\text{Mpc/h} \leq z \leq 1010\text{Mpc/h}$. Here the persistence threshold given to DisPerSE is varied between no cuts (a), $\sigma = 1$, $\sigma = 2$ and $\sigma = 3$. Each blue dot represents a halo particle while all lines represents filaments assigned by DisPerSE.	65
6.3	Figure showing histograms of filament lengths calculated by disperse for the MD1 dataset. In this figure, the persistence threshold given to DisPerSE is varied between no cuts (a), $\sigma = 1$ (b), $\sigma = 2$ (c), $\sigma = 3$ (d).	66
6.4	Figure showing histograms of filament lengths calculated by disperse the MD2 dataset. In this figure, the persistence threshold given to DisPerSE is varied between no cuts (a), $\sigma = 1$ (b), $\sigma = 2$ (c), $\sigma = 3$ (d).	67
6.5	Figure showing the calculated density profile $\rho_g(r)/\bar{\rho}$ for filaments in the MD1 dataset with $\sigma = 1$	69
6.6	Figure showing the calculated density profile $\rho_g(r)/\bar{\rho}$ for filaments in the MD1 dataset with $\sigma = 2$	70
6.7	Figure showing the calculated density profile $\rho_g(r)/\bar{\rho}$ for filaments in the MD2 dataset with $\sigma = 1$	71
6.8	Figure showing the calculated density profile $\rho_g(r)/\bar{\rho}$ for filaments in the MD2 dataset with $\sigma = 2$	72

6.9	Figure showing the calculated radial velocity profile compared with the predictions from linear theory for filaments in the MD1 dataset with $\sigma = 1$.	74
6.10	Figure showing the calculated radial velocity profile compared with the predictions from linear theory for filaments in the MD1 dataset with $\sigma = 2$.	75
6.11	Figure showing the calculated radial velocity profile compared with the predictions from linear theory for filaments in the MD2 dataset with $\sigma = 1$.	76
6.12	Figure showing the calculated radial velocity profile compared with the predictions from linear theory for filaments in the MD2 dataset with $\sigma = 2$.	77
6.13	Figure showing a histogram of the effective void radius of voids found in the Multidark 2 dataset.	78
6.14	Figure showing a histogram of the effective void radius of voids found in the Multidark 3 dataset.	79
6.15	Figure showing a histogram of the effective void radius of voids found in the Multidark 4 dataset.	80
6.16	Figure showing the density profile for voids found in the Multidark 2 dataset.	81
6.17	Figure showing the density profile for voids found in the Multidark 3 dataset. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ in the bottom left, and $r \geq 40\text{Mpc/h}$ in the bottom right.	82
6.18	Figure showing the density profile for voids found in the Multidark 4 dataset. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ in the bottom left, and $r \geq 40\text{Mpc/h}$ in the bottom right.	83
6.19	Figure showing the calibrated dark matter density profile, in accordance with the method described in section 5.6, for voids found in the Multidark 2 dataset. The blue line is the dark matter density profile multiplied by a factor 1.85 in order to make it resemble $\delta_g(r)$. The orange line is the scaled dark matter density profile divided by the bias factor for the given MultiDark dataset. The orange line is the one used in the calculation of the model.	84
6.20	Figure showing the calibrated dark matter density profile, in accordance with the method described in section 5.6, for voids found in the Multidark 3 dataset. The blue line is the dark matter density profile multiplied by a factor 1.85 in order to make it resemble $\delta_g(r)$. The orange line is the scaled dark matter density profile divided by the bias factor for the given MultiDark dataset. The orange line is the one used in the calculation of the model.	85

6.21	Figure showing the calibrated dark matter density profile, in accordance with the method described in section 5.6, for voids found in the Multidark 4 dataset. The blue line is the dark matter density profile multiplied by a factor 1.85 in order to make it resemble $\delta_g(r)$. The orange line is the scaled dark matter density profile divided by the bias factor for the given MultiDark dataset. The orange line is the one used in the calculation of the model.	86
6.22	Figure showing the density contrast $\Delta_v(r)$ for voids found in the Multidark 2 dataset.	87
6.23	Figure showing the density contrast $\Delta_v(r)$ for voids found in the Multidark 3 dataset. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ in the bottom left, and $r \geq 40\text{Mpc}/h$ in the bottom right.	88
6.24	Figure showing the density contrast $\Delta_v(r)$ for voids found in the Multidark 4 dataset. Here the radius of voids considered are in the radius range of: all voids in top figure, $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ in the bottom left, and $r \geq 40\text{Mpc}/h$ in the bottom right.	89
6.25	Figure showing the radial velocity profile of halos around voids found in the Multidark 2 dataset. Here it is compared with the prediction from linear theory, given in equation 5.12, as well as the model with the correction term, given in equation 5.13.	91
6.26	Figure showing the radial velocity profile of halos around voids found in the Multidark 3 dataset. Here it is compared with the prediction from linear theory, given in equation 5.12, as well as the model with the correction term, given in equation 5.13. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ in the bottom left, and $r \geq 40\text{Mpc}/h$ in the bottom right.	92
6.27	Figure showing the radial velocity profile halos around voids found in the Multidark 4 dataset. Here it is compared with the prediction from linear theory, given in equation 5.12, as well as the model with the correction term, given in equation 5.13. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ in the bottom left, and $r \geq 40\text{Mpc}/h$ in the bottom right.	93
6.28	Figure showing the velocity dispersion, given by equation 5.23, for halos around voids found in the Multidark 2 dataset.	94
6.29	Figure showing the velocity dispersion, given by equation 5.23, for halos around voids found in the Multidark 3 dataset. In this figure, the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ in the bottom left, and $r \geq 40\text{Mpc}/h$ in the bottom right.	95

6.30	Figure showing the velocity dispersion, given by equation 5.23, for halos around voids found in the Multidark 4 dataset. In this figure, the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ in the bottom left, and $r \geq 40\text{Mpc}/h$ in the bottom right.	96
6.31	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 2 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. The fiducial value for each parameter is overplotted.	99
6.32	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. The fiducial value for each parameter is overplotted.	100
6.33	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. In this figure, only voids with effective radius $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ are considered. The fiducial value for each parameter is overplotted.	101
6.34	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. In this figure, only voids with effective radius $\geq 40\text{Mpc}/h$ are considered. The fiducial value for each parameter is overplotted.	102
6.35	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory given by equation 5.12. The fiducial value for each parameter is overplotted.	103
6.36	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. In this figure, only voids with effective radius $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ are considered. The fiducial value for each parameter is overplotted.	104
6.37	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. In this figure, only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered. The fiducial value for each parameter is overplotted.	105
6.38	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 2 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13.	108

6.39	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13.	109
6.40	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13. In this figure, only voids with effective radius $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ are considered.	110
6.41	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13. In this figure only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered.	111
6.42	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13.	112
6.43	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13. In this figure only voids with effective radius $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ are considered.	113
6.44	Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13. In this figure only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered.	114
6.45	Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 2 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given in equation 5.12	117
6.46	Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 3 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given by equation 5.12	118
6.47	Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 3 dataset using a dark matter profile ,as described in section 5.6, and modelling the velocity using linear theory given in equation 5.12. In this figure only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered.	119
6.48	Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 4 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given in equation 5.12	120

6.49	Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 4 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given in equation 5.12. In this figure only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered.	121
6.50	Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD2 dataset.	124
6.51	Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD3 dataset.	125
6.52	Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD3 dataset. Here only voids with effective radius $r \geq 40\text{Mpc}/h$ is included.	126
6.53	Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD4 dataset.	127
6.54	Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD4 dataset. Here only voids with effective radius $r \geq 40\text{Mpc}/h$ is included.	128

Chapter 1

Introduction

The current dominating model for the universe is the Λ CDM model. This model assumes that the mass-energy density of the universe is made up of three major components, namely dark energy Λ , cold dark matter CDM and regular baryonic matter. The modeling of the evolution and expansion of the universe is related to the mass-energy density of the universe itself. An important aspect of cosmology is therefore to constrain the bounds on these components in order to figure out how large the fraction of the total mass-energy density of the universe each of them makes up.

The leading experiments related to constraining cosmological parameters is the study of radiation remnant from the very early universe, namely the cosmological microwave background (CMB)[5], with the COBE-[6] WMAP-[7] and Planck-satellites[8] in the forefront. Other studies include that of stellar supernova type-Ia explosions[9]. Advances in cosmological experiments over the last decades have brought cosmology into the era of "precision cosmology", now that cosmological parameters can be constrained with sub-percent accuracy. This has led to tensions in that certain experiments may disagree over values for cosmological parameters. An example of this is the Hubble tension in which measurements of the CMB disagree with that of type Ia supernovae when measuring the evolution of the Hubble parameter, which quantifies the expansion of the universe. This particular tension could be a case of statistical fluctuation or systematic error, but these situations do warrant more independent experiments in order to strengthen the current knowledge.

Additional independent sets of experiments relating cosmological components to observables can be found through studies of the distribution of matter in the universe, namely the large-scale structure (LSS) of the universe. Studies of the large-scale structure of the universe include the study of how matter is clustered at different cosmological redshifts in order to determine statistical properties of the matter distribution. On large scales, the matter distribution in the universe forms a web-like pattern known as the cosmic web[10]. Substructures of the cosmic web include voids, empty regions of space containing little to no matter. These voids are bound by structures known as filaments. Filaments consist of gravitationally bound galaxy clusters.

One of the main statistical properties of the LSS that one can measure is the angular correlation function between definable physical objects. In short, a correlation function in this context measures the excess probability over random that two objects are separated by a distance r . The imprints of baryonic acoustic oscillations on the correlation between galaxy pairs[11], which appear as ripples in the density distribution of baryonic matter, serve as a standard ruler when measuring cosmological parameters. The effects this has on the correlation function between matter has been studied both theoretically[12] and through observational experiments[13]. In addition, other studies like [14] and [15] show that the statistical properties of galaxy clustering provides useful constraints on cosmological parameters. Another statistical tool that allows for the extraction of information from the density field is the void-galaxy cross-correlation function. The fact that the universe is expanding[16], causes light travelling towards us to be redshifted. With current observational experiments, the only reliable way to observe the LSS is through observations of electromagnetic radiation. When observing objects in the universe, the redshifted light causes the observed matter distribution to appear distorted along the line of sight direction. This introduces anisotropy in the void-galaxy correlation function. Work on modelling the anisotropic effects on the void-galaxy correlation function as a result of redshift space distortions has been conducted [17]. Additional works like [18] and [3] have shown that this provides additional constraints to cosmological parameters. These independent experiments allow one to extract even more information about underlying cosmology from the density field.

The filaments themselves have not yet received the proper treatment as that of voids and galaxy clusters. Although great efforts have been made analysing many of their properties[19], their matter distribution has not yet received a thorough statistical review as that of voids and the matter distribution in general. Further contribution to this field includes proper modelling of the physical behavior, such as the velocity field of matter around filaments. By modelling the statistical properties of filaments correctly, one can potentially gain an additional set of experiments for extracting cosmological properties from the density field. This can potentially serve as a useful supplement for the studies conducted on voids and galaxy clustering. Additionally, future galaxy surveys are expected to greatly increase the size of observational datasets. Current surveys of the large scale structure of the universe include the Sloan Digital Sky Survey [20], in which the redshifts of 1.5 million galaxies is measured. Future surveys include the Euclid mission, in which one of the scientific goals is to measure the redshifts upwards of 50 million galaxies[21]. In preparation for future experiments like Euclid, which will greatly improve the data for which LSS studies are to be carried out, simulation data is useful for testing and developing methodology. Current computer simulations enable the study of the evolution of billions of dark matter particles[2][22], which allows for relatively accurate representations of the real universe. Since future cosmological surveys and experiments allow for even more precise measurements, additional emphasis on mapping the potential error of applied models is crucial. This is in order to fully extract as much knowledge as possible from future surveys and experiments. As mentioned, the correlation function for voids and galaxies has been studied in previous works. How-

ever, in preparation for future observational experiments, applying methodology and potential improvements on large datasets provide useful testing.

The work in this thesis is divided into two main parts. Using data from The Big MultiDark Planck simulation (BigMDPL)[22], the first part of this work contains an analysis of filaments where the goal is to test a model for the velocity component of dark matter halos perpendicular to the filament spine. Using linear perturbation theory, I will derive a simple model for the radial velocity perpendicular to the spine of cosmological filaments. This model will be compared to velocity data from numerical simulations in order to test whether this is a valid approximation for it to be used in further modelling of the statistical properties of filaments. The second part of this thesis includes a test of the void-galaxy correlation function applied on data from The Big MultiDark Planck simulation (BigMDPL)[22]. The methodology will be tested using three different samples with a different number of dark matter halos ranging from 3.5 million to 15 million. In addition, proposed improvements to the modelling of velocities of matter around voids[4] will be tested in order to see if this will improve the performance of the models.

This thesis is divided into multiple sections. Chapter 2 provides a summary of cosmological theory used in the development of the methodology. It also provides a short summary of the fundamental statistical theory used and cosmological simulations in general. Chapter 3 provides an introduction to the codes utilized for identifying filaments and voids from the simulated distribution of dark matter halos. This is followed by a short introduction of the dataset and its properties given in chapter 4. The methodology particular to the work conducted in this thesis itself is introduced in chapter 5. This is followed chapter 6 which presents the results and a discussion of their properties. This section is split into two parts. Section 6.1 provides the results and a discussion for the work done on filaments, while section 6.2 provides the same for the work done on voids. Finally, some concluding remarks are given in chapter 7.

Chapter 2

Background Theory

2.1 History of cosmology and how we model the universe

This section aims to introduce much of the cosmology relevant to the methodology and analysis applied in this thesis. The following subsections are largely based on [23], [24] and [25] unless otherwise noted.

2.1.1 Distances in the universe and cosmological redshift

Distances in an expanding universe can prove hard to fathom. In our everyday lives, we are used to measuring distances in meters and kilometers. The average distance from the earth to the sun however is approximately $1.49 \cdot 10^{11}$ m. Although this distance on a cosmic scale is very short, the numbers in our everyday units of measurement already start to become too big for us to have a reasonable idea of how long this distance actually is. We have a clear idea of what a meter and a kilometer is as we have to deal with them every day. To make things more manageable when measuring these large distances, the average distance from the earth to the sun has been assigned its own unit of distance measurement known as the Astronomical unit (AU). This unit is a useful tool for measuring relative distances in our own solar system, but the star nearest to our solar system, Proxima Centauri, is located at approximately 268394AU away from our closest star the Sun. Therefore it is convenient to introduce another unit of distance measurement known as a light-year. This unit with a value of $9.4607 \cdot 10^{15}$ m is the distance at which light travels in vacuum during one Julian year equal to 365.25 days. This gives us a more manageable account of the distance from the Sun to Proxima Centauri as 4.244 light-years [26]. With both the AU and the light-year introduced we can define the last unit for distance measurement, namely the parsec equal to 3.26 light-years. One parsec is equal to the distance at which 1 AU subtends an angle of one arcsecond. Most cosmological distances are measured in parsec, abbreviated Pc, with either the prefix mega or giga as cosmology deals mostly with intergalactic scales much larger than the average distance between two galaxies.

Due to the fact that the universe is expanding, light travelling towards us from far

away objects gets redshifted. Since the expansion of the universe is isotropic, i.e, the expansion rate of the universe is equal in all directions, the redshift of this light can be used to measure the distance at which we observe objects in the universe. The cosmic redshift z is given by the relation

$$z = \frac{\lambda_{obs}}{\lambda_{em}} - 1, \quad (2.1)$$

where λ_{obs} is the observed wavelength we measure here on earth and λ_{em} is the wavelength emitted by the observed object. By combining this with the doppler formula $\frac{\lambda_{obs}-\lambda_{em}}{\lambda_{em}} = \frac{v}{c}$, one gets

$$z = \frac{v}{c}, \quad (2.2)$$

where v is the velocity of the observed object relative to the observer. The velocity of an object due to the expansion of the universe is given by Hubble's law

$$v = H_0 d_p, \quad (2.3)$$

where d_p is what is known as proper distance and H_0 is the Hubble parameter at present time. The Hubble parameter measures the expansion rate of the universe as a function of time. The proper distance can be interpreted as the length measured if one was to use a ruler between two objects a and b at a fixed time t taking into account the fact that our universe is expanding. Before introducing how proper distance is calculated, it can be useful to talk about the concept of comoving distance first. If we choose a time $t = t_0$, which is present time and introduce a coordinate system at this time, we can choose an arbitrary point to center ourselves at in this coordinate system. If we apply a ruler in this coordinate system and measure the distance from our point to another arbitrary point, this is the comoving distance r . The comoving distance is simply the distance measured in a coordinate system at $t = t_0$. The proper distance on the other hand factors in the expansion of the universe through the scale factor $a(t)$. The proper distance is given as

$$d_p = \int_{r_0}^{r_1} a(t) r dr, \quad (2.4)$$

where $a(t)$ is a quantity known as the scale factor. The scale factor parametrizes the relative expansion of the universe. It is defined in such a way that a_0 , subscript 0 meaning $a(t = t_0)$, is equal to one. This definition of proper distance also implies that comoving distance is the proper distance at $t = t_0$. With these properties in place we can now see how redshift, which is a measurable quantity, can be used to measure the real distance to objects in our universe.

Another measurable quantity is flux emitted from an astronomical object. The flux F is defined as

$$F = \frac{L}{4\pi R^2}, \quad (2.5)$$

where L is the luminosity and R is the distance to the object. Luminosity is a measure of the absolute electromagnetic power emitted from an astronomical object. The flux,

however, measured in W/m^2 , is simply the luminosity divided by the area of a shell with a radius at which we observe the object, assuming the radiation is isotropically emitted. Some astronomical phenomena, like a type 1a supernova, have a fixed luminosity due to the characteristics of the process that creates it. This is called a standard candle. By measuring the observed flux from such a phenomenon, one can then calculate the luminosity distance as

$$d_L = \sqrt{\frac{L}{4\pi F}}. \quad (2.6)$$

Lastly, we introduce the angular diameter distance. The angular extent of an object in the sky provides us with another measurable quantity related to distances in the universe. The angular diameter distance is, for an object with a known size D at a large distance d , where $d \gg D$, given as

$$d_A \equiv \frac{D}{\Delta\theta}. \quad (2.7)$$

Here $\Delta\theta$ is the angle covered by the object measured in radians and d_P is the proper distance. Taking into account the proper distance separating the two endpoints of the diameter of the observed object we have

$$D = a(t)r\Delta\theta, \quad (2.8)$$

where r is the distance from the observer to the measured object. This will give an expression for the angular diameter distance

$$d_A = \frac{r}{1+z}, \quad (2.9)$$

where the relation $a(t) = 1/(1+z)$ has been used.

2.1.2 Modelling the evolution of the universe and the Λ CDM model.

The cosmological principle states that on sufficiently large scales the universe is isotropic and homogenous. This means the universe looks the same in every direction for every observer at any point in space for sufficiently large scales. By sufficiently large scales one means distances roughly 100Mpc or larger [25, p. 12]. This has proved to be a good assumption by surveys like the Sloan Digital Sky Survey[27]. Figure 2.1 shows a sky map from a survey released by the SDSS project.

The fact that the universe was homogeneous and isotropic was an essential assumption for a particular solution of the Einstein field equation. This particular equation, first published by Albert Einstein in 1915 [28] concerning his theory of general relativity, is given as

$$G_{\mu\nu} = 8\pi GT_{\mu\nu}. \quad (2.10)$$

This equation relates $G_{\mu\nu}$, which is the Einstein tensor in which describes the curvature of spacetime, with the energy momentum-tensor $T_{\mu\nu}$, which describes the energy content of the universe. During the 1920s, by assuming the universe is isotropic, homogenous

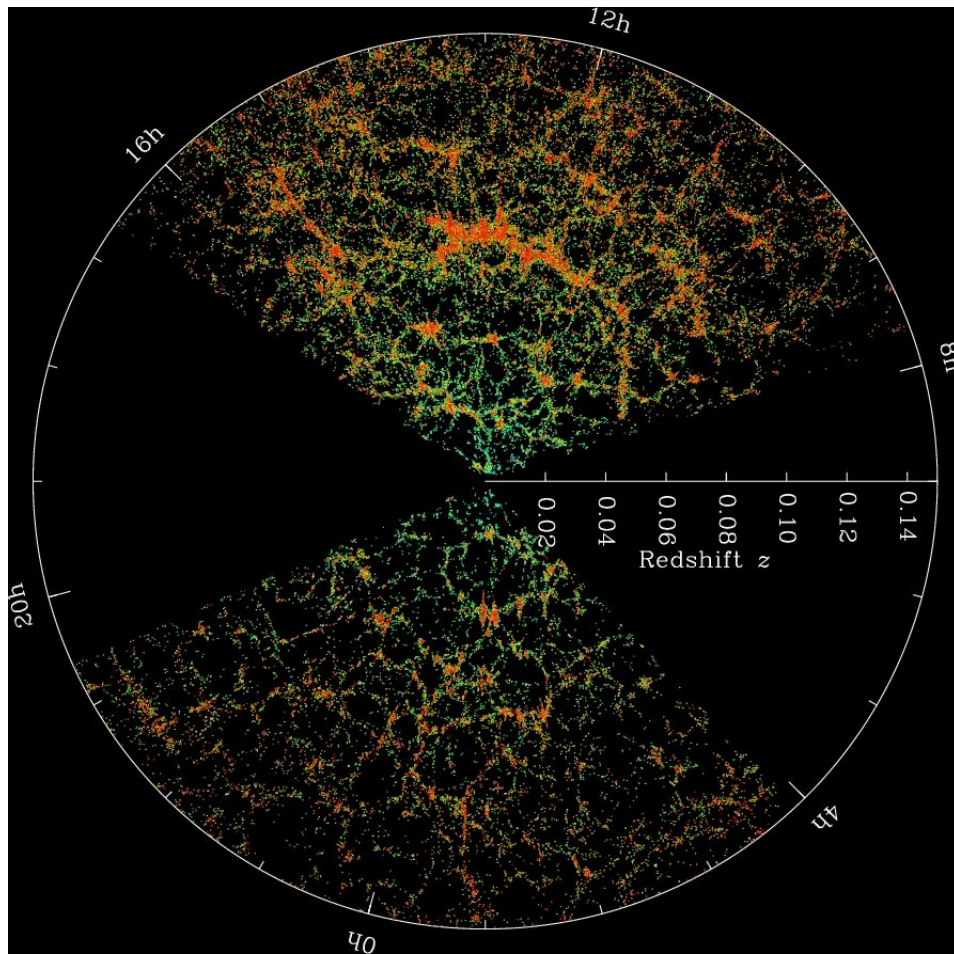


Figure 2.1: The SDSS (Sloan Digital Sky survey) map of the universe. Each dot represents a galaxy and the color scale represents the local density. Downloaded from the SDSS website <https://www.sdss.org/science/orangepie/>. Image Credit: M. Blanton and SDSS.

and expanding, the scientists Alexander Friedmann and Georges Lemaître worked on finding a solution to equation 2.10. In the 1930s Howard P. Robertson and Arthur Geoffrey Walker further explored the problem and proved that this is the only solution on a spacetime that is homogenous and isotropic. This solution, named after the four scientists, is known as the Friedman-Lemaître-Robertson-Walker metric. The metric is a tool that allows us to calculate the physical distance in a coordinate system. For example, for the Cartesian coordinate system, we have the metric

$$g_{ij} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (2.11)$$

The definition of the line element ds^2 in this case, is given as

$$ds^2 = \sum_{i,j=0}^1 g_{ij} dq^i dq^j, \quad (2.12)$$

where $q = (x, y)$ is a vector representing the spatial dimensions and i and j represents the indices. This gives the familiar line element $ds^2 = dx^2 + dy^2$. For a flat universe the Friedman-Lemaître-Robertson-Walker metric for four dimensional space time takes the following form

$$g_{\mu\nu} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -a^2 & 0 & 0 \\ 0 & 0 & -a^2 & 0 \\ 0 & 0 & 0 & -a^2 \end{bmatrix}. \quad (2.13)$$

The line element is now given as

$$ds^2 = \sum_{\mu,\nu=0}^3 g_{\mu\nu} dx^\mu dx^\nu, \quad (2.14)$$

where $dx^0 = dt$ and the remaining three indices are for the three spatial coordinates. This gives us the following line element

$$ds^2 = dt^2 - a^2(dx^2 + dy^2 + dz^2). \quad (2.15)$$

The energy momentum tensor $T_{\mu\nu}$ for a perfect isotropic fluid takes the form

$$T_{\mu\nu} = \begin{bmatrix} -\rho & 0 & 0 & 0 \\ 0 & -a^2 P & 0 & 0 \\ 0 & 0 & -a^2 P & 0 \\ 0 & 0 & 0 & -a^2 P \end{bmatrix}, \quad (2.16)$$

where P is the pressure. Using this metric together with the Einstein field equations results in the important equations known as the first and second Friedman equations respectively as

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} \quad (2.17)$$

and

$$\frac{\ddot{a}^2}{a^2} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right). \quad (2.18)$$

I refer the reader to [24, ch. 2] for an explanation on how to derive the Friedmann equations starting with the Friedman-Lemaître-Robertson-Walker metric. These differential equations relate the evolution of the scale factor a with the following properties of the universe, namely density ρ , pressure p and gravitational constant G . In these equations \dot{a} and \ddot{a} denotes derivative and double derivative with respect to time, meaning that $a = a(t)$ is a function of time.

In modern cosmology, the standard model for describing the universe is the Λ CDM model. Its name is an abbreviation for what is considered the main energy contributions governing the expansion of the universe. The cosmological constant Λ represents dark energy, CDM is an abbreviation for cold dark matter and lastly, we have ordinary baryonic matter, which is what we interact with in our everyday lives. The current estimates suggest that approximately 69% of the universe consists of dark energy while the remaining 31% is attributed to matter. The matter contribution is made up of 27% dark matter and 4% is regular baryonic matter. This leaves only trace contributions from other energy contributing factors such as photons (γ) and neutrinos (ν) [29]. When modelling the universe these quantities enter into what is called density parameters Ω_i , where i represents a certain type of energy contribution to the universe i.e dark matter or CDM. The density parameter is defined as

$$\Omega_{i,0} = \frac{\rho_{i,0}}{\rho_{c,0}}, \quad (2.19)$$

where $\rho_{i,0}$ is the density of a given energy contribution and $\rho_{c,0}$ is the critical density of the universe. Subscript zero indicating these quantities measured at present time. The critical density is the density at which the universe is flat. The critical density at a given time is given as $\rho_c = \frac{3H(t)^2}{8\pi G}$. For a flat universe, we have

$$\sum_i \Omega_i = 1. \quad (2.20)$$

By combining both Friedmann equations (equations 2.17 and 2.18), one can derive the following relation for the density ρ

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right). \quad (2.21)$$

By inserting the equation of state $p = w\rho c^2$, where w is a constant, into equation 2.21 one will get the following differential equation

$$\dot{\rho} = -3\frac{\dot{a}}{a}(1+w)\rho. \quad (2.22)$$

This equation can be integrated for an arbitrary a and ρ to present time giving

$$\rho_i = \rho_{i,0}a^{-3(1+w_i)}, \quad (2.23)$$

where subscript i is again added to denote a specific energy contribution to the universe. The constant w_i takes different values when considering the different energy contributions. For CDM and baryons we have $w = 0$, for radiation $w = \frac{1}{3}$ and for the cosmological constant we have $w = -1$. By taking into consideration equation 2.17 and dividing both sides by H_0^2 , and recognize the critical density $\rho_{c,0} = \frac{3H_0^2}{8\pi G}$, one can rewrite the first Friedmann equation as

$$\frac{H(t)^2}{H_0^2} = \frac{\rho}{\rho_{c,0}}. \quad (2.24)$$

By separating the density into its different components as $\rho = \sum_i \rho_i$ and using equation 2.23 for the different components, one can rewrite the first Friedmann equation using equation 2.19 as

$$H(t) = H_0\sqrt{(\Omega_{CDM,0} + \Omega_{b,0})a^{-3} + \Omega_{r,0}a^{-4} + \Omega_{\Lambda,0}}. \quad (2.25)$$

Where b represents baryonic matter and r represents radiation.

2.1.3 Einstein-de Sitter model

While the Λ CDM model is the leading model for describing the evolution of our universe, other universe models, being simpler to work with, will also prove beneficial as approximations for different applications. One of these models is the Einstein-de Sitter model [30] proposed by Albert Einstein and Willem de Sitter in 1932. This model contains only matter making $\Omega_m = 1$. After learning of the expansion of the universe, from the observations of Edwin Hubble, Einstein removed the cosmological constant from his equation as this constant was first proposed to keep the universe static. After the discovery that the expansion of the universe is accelerating [31][32] the cosmological constant was reintroduced. While the Einstein-de Sitter model is long discarded, it is still useful as it is easy to work with in an analytical framework and provides good approximations for different applications.

In a universe containing only matter, the first Friedmann equation with $w = 0$ takes the form

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho_{m,0}a^{-3}. \quad (2.26)$$

By multiplying with H_0^2/H_0^2 and remembering that $\rho_{c,0} = \frac{3H_0^2}{8\pi G}$, one gets

$$a\dot{a}^2 = H_0^2\frac{\rho_{m,0}}{\rho_{c,0}}. \quad (2.27)$$

Since we consider a flat universe containing only matter we have $\frac{\rho_{m,0}}{\rho_{c,0}} = \Omega_{m,0} = 1$. Taking the square root of both sides, we get

$$a^{1/2}\dot{a} = H_0. \quad (2.28)$$

This gives us a differential equation which can be integrated as

$$\int_{a_0}^a a^{1/2} da = H_0 \int_{t_0}^t dt, \quad (2.29)$$

which gives the solution

$$\frac{2}{3}(a^{3/2} - a_0^{3/2}) = H_0(t - t_0). \quad (2.30)$$

By choosing time $t = 0$, in which $a = 0$, and using that $a_0 = 1$, we get

$$\frac{2}{3} = H_0 t_0. \quad (2.31)$$

By imposing this boundary condition we find that

$$a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3}}. \quad (2.32)$$

2.1.4 Growth of matter perturbations in linear perturbation theory.

As I have previously stated, observational data suggest that the universe is isotropic and homogenous. This however was for sufficiently large scales at around 100 Mpc or larger. On smaller scales, the universe is not isotropic and homogenous. A single galaxy, for example, is denser than the intergalactic medium that separates it from other galaxies. The anisotropy in temperature of the cosmic microwave background (CMB), as measured by the Planck satellite and its predecessor CMB experiments, is measured to be around $\Delta T/T = 10^{-5}$. This shows that at around redshift $z = 1000$, we had relatively small fluctuations while we today observe galaxy clusters which has around 200 times larger density than the average density of an equal sphere in the universe [23, p. 342]. This suggests that the original density perturbations grows over time. The small structure deviations at early times evolve through what is known as gravitational instability. Regions with higher density will attract more material through gravity giving a flow of matter towards dense regions making them even denser. This makes an already irregular distribution of matter become even more irregular with time and will evolve from the small fluctuations at the time of the CMB to the dense structures we observe in our universe today. To describe this evolution, one defines the relative density contrast as

$$\delta(\mathbf{r}, t) \equiv \frac{\rho(\mathbf{r}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}. \quad (2.33)$$

Here $\bar{\rho}(t)$ denotes the mean matter density of the whole universe at time t , while $\rho(\mathbf{r}, t)$ denotes the local density at position \mathbf{r} at time t .

The growth of these perturbations can, on scales substantially smaller than the Hubble radius, which is the radius of the observable universe, be described in the framework of linear perturbation theory. On these scales, Newtonian gravity is sufficient to describe the nature of structure growth. With the approximation that the universe only consists of pressureless matter, which is the case for cold dark matter, which is the main constituent of the matter in the universe, the matter distribution of the universe is modelled as a pressureless fluid. The equations of motion for such a fluid is governed by the equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (2.34)$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla P}{\rho} - \nabla \Phi \quad (2.35)$$

and

$$\nabla^2 \Phi = 4\pi G \rho - \Lambda. \quad (2.36)$$

Here $\mathbf{v} = \mathbf{v}(\mathbf{r}, t)$ is the velocity field of the fluid and $\Phi = \Phi(\mathbf{r}, t)$ is the newtonian gravitational potential. P is the pressure of the fluid. Equation 2.34 is the continuity equation which tells us that the density changes with the flow of mass. Equation 2.35, which is the Euler equation, describes the behavior of the fluid under the influence of an external force. The last equation, equation 2.36, is the Poisson equation which relates the gravitational potential to the density field. These equations are solvable in the limit $|\delta| \ll 1$, and can then be used to derive an expression for relatively small density contrasts, which can be used to describe the evolution of density perturbations on large scales. We can model small perturbations to first order by substituting $\rho = \rho_0 + \delta\rho$, $\mathbf{v} = \mathbf{v}_0 + \delta\mathbf{v}$, $P = P_0 + \delta P$ and $\Phi = \Phi_0 + \delta\Phi$. Here δ is used to assign small perturbations to the physical quantities and should not be confused with the previously defined density contrast. The velocity also has to take into account the expansion of the universe in addition to the peculiar velocity of the particles giving $\mathbf{v} = H(t)\mathbf{r} + \mathbf{v}_{pec}$. These equations can then be reduced to

$$\frac{\partial^2 \delta}{\partial t^2} + 2H(t) \frac{d\delta}{dt} = 4\pi G \bar{\rho} \delta. \quad (2.37)$$

[23, p. 345]. This equation only contains time derivatives, and none of the coefficients is dependent on \mathbf{x} . Therefore $\delta(\mathbf{x}, t)$ can then be expressed as a spatial and time-dependent expression on the form

$$\delta(\mathbf{x}, t) = D(t)g(\mathbf{x}). \quad (2.38)$$

Here $g(\mathbf{x})$ is an arbitrary function of the spatial comoving coordinate \mathbf{x} , and $D(t)$ satisfies

$$\frac{\partial^2 D}{\partial t^2} + 2H(t) \frac{dD}{dt} = 4\pi G \bar{\rho} D. \quad (2.39)$$

This equation has two solutions. One solution is strictly increasing and one is strictly decreasing. As time evolves the decreasing solution will become negligible and the increasing solution will dominate. This solution, denoted as $D_+(t)$, is called the growth

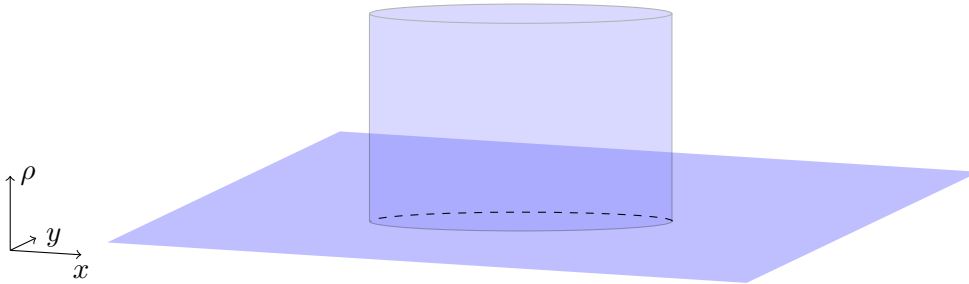


Figure 2.2: A simple illustration of the spherical top hat model.

factor and determines the amplitude of structure growth as a function of time. Since $t \propto a$, one can then model the growth factor for a given cosmological model. We will lastly define the function

$$f(a) \equiv \frac{d \log D_+}{d \log a}, \quad (2.40)$$

to quantify the relationship between the growth of structure and the expansion of the universe.

2.1.5 Spherical collapse models

Linear perturbation theory has its limitations. In particular, it will not be able to accurately describe the formation of dark matter halos or galaxy clusters. The evolution of density contrasts of order higher than unity requires different techniques to model in a reasonable way. The spherical top-hat model is one of the most fundamental frameworks to analytically model the non-linear collapse of overdensities in the universe. The spherical top-hat model gets its name from how its density profile is defined. For the spherical top-hat model we have

$$\rho(\mathbf{r}, t) = \begin{cases} \bar{\rho}(1 + \delta_i), & |\mathbf{r} - \mathbf{r}_c| \leq R \\ \bar{\rho}, & |\mathbf{r} - \mathbf{r}_c| > R, \end{cases} \quad (2.41)$$

where \mathbf{r}_c is the center of the overdensity and R is the radius. δ_i is a factor determining the size of a given overdensity. Figure 2.2 gives a simple graphical illustration of the model, hence why it has the name spherical top-hat model. For a spherical shell with the given density we have the mass

$$M = \frac{4}{3} \pi R^3 \bar{\rho}(1 + \delta_i). \quad (2.42)$$

The evolution of the radius R can be modelled through Newtons second law as

$$\frac{d^2 R}{dt^2} = -\frac{GM}{R^2}. \quad (2.43)$$

By multiplying both sides of this equation with $\frac{dR}{dt}$ and integrating one gets

$$\frac{1}{2} \frac{d^2 R}{dt^2} - \frac{GM}{R} = E. \quad (2.44)$$

Here it is worth noting that the left-hand side of 2.44 is the sum of the kinetic and potential energy. Therefore the integration constant E can be interpreted as the total energy. It can be shown that for $E < 0$, this equation has the parametrised solution

$$R = A(1 - \cos\theta) \quad (2.45)$$

$$t = B(\theta - \sin\theta) \quad (2.46)$$

$$A^3 = GMB^2. \quad (2.47)$$

[12, p. 79]. By setting an initial radius R_i at time t_i this solution tells us that a shell expanding with the background evolution of the universe will eventually slow down and reach a maximum radius at $\theta = \pi$. This is called turn-around. At this point, the sphere will begin to collapse. This solution also suggests that at $\theta = 2\pi$ the sphere would collapse to a point. This is not the case as something known as virialization occurs before this happens.

The virial theorem describes the relationship between the potential and kinetic energy in dynamical equilibrium in a set of collisionless particles under the influence of a gravitational potential. Individual particles in the gas will reach $R = 0$ and eventually shoot outwards again. The whole system will stabilize when it has virialized. The virial theorem states that

$$|U| = 2K, \quad (2.48)$$

where U is the potential energy and K is the kinetic energy. At the radius of turn around, the kinetic energy is zero. As the cloud collapses the kinetic energy will increase and eventually it will virialize when the conditions of equation 2.48 is fulfilled.

2.1.6 Two point correlation functions and the matter power spectrum.

Matter in the universe is not randomly distributed in a uniform fashion. Starting with the initial perturbations, matter clumped together and formed galaxies and clusters as a result of gravitational instability. This means that for a given random point in space, the probability of finding a galaxy in the vicinity of that point is not independent of the surrounding galaxy distribution. The statistical distribution of galaxies is described by the two-point correlation function. The two-point correlation function for galaxies $\xi_g(r)$ describes the probability of two galaxies are separated by a distance r . The correlation function is related to the density contrast as

$$\xi(|\mathbf{r}_1 - \mathbf{r}_2|) = \langle \delta(\mathbf{r}_1) \delta(\mathbf{r}_2) \rangle. \quad (2.49)$$

The matter power spectrum $P(k)$ provides a statistical description of the distribution of matter in the universe, but in Fourier space. Figure 2.3 shows the matter power spectrum inferred from a multitude of cosmological surveys. The matter power spectrum

describes the amplitude of the density contrast on different length scales $L = \frac{2\pi}{k}$, where k is the Fourier wave number. The matter power spectrum is related to the two-point correlation function as a Fourier transform given by

$$P(k) = 2\pi \int_0^\infty x^2 \frac{\sin(kx)}{kx} \xi(x) dx. \quad (2.50)$$

Both of these quantities are statistical properties and are of high importance when

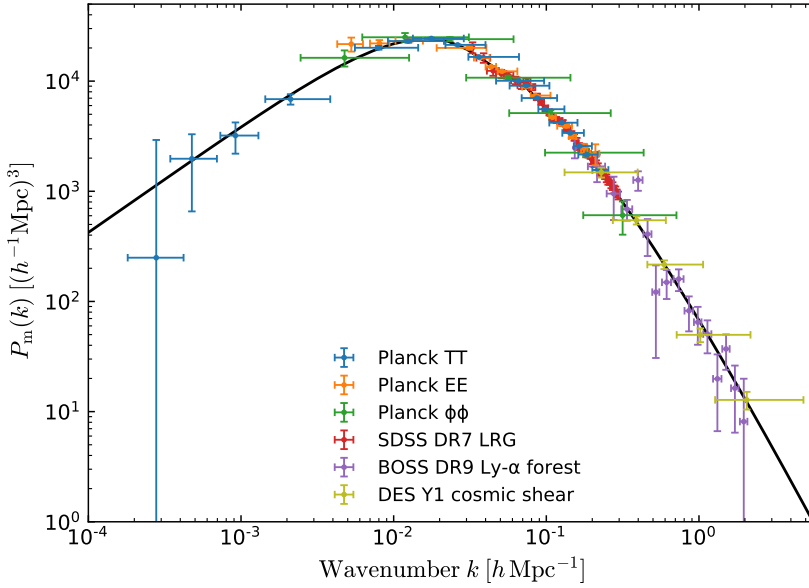


Figure 2.3: The matter power spectrum inferred from different cosmological surveys. Image credit: "ESA and the Planck Collaboration" [1].

comparing models to observations or simulations. The initial conditions for the origin of structure in our universe are coupled to the period of Inflation. Inflation is a rapid expansion of space that occurred in the infancy of the universe after The Big Bang. Inflationary theory does not predict the exact location of the distributed matter. Instead, it predicts how it should be statistically distributed. A simulation or calculation of structure growth may not look identical to anything anywhere in the universe (if it is not infinitely large), but if the statistical properties are the same, then the models are correct. It is therefore important to note that two universes are considered identical if they inhabit the same statistical properties. This makes the power spectrum and two-point correlation function important vital tools when studying the universe.

An important cosmological parameter is σ_8 which serves as a normalization for the matter power spectrum. The parameter σ_8 is motivated by the observation that when averaging over spheres on scales of radius $R = 8\text{Mpc}/h$ it is found that

$$\sigma_{8,g}^2 \equiv \left\langle \left(\frac{n - \bar{n}}{\bar{n}} \right)^2 \right\rangle_{8\text{Mpc}/h} \approx 1, \quad (2.51)$$

where n is the local number density of galaxies and \bar{n} is the average number density. In a similar fashion, it is defined through the dark matter density contrast as

$$\sigma_8^2 = \left\langle \left| \delta_{8\text{Mpc}/h} \right|^2 \right\rangle, \quad (2.52)$$

where the density contrast δ is averaged over spheres of $R = 8\text{Mpc}/h$. These two parameters can be related through the linear bias factor as

$$\sigma_8 = \frac{\sigma_{8,g}}{b}. \quad (2.53)$$

Since the overdensity of galaxies does not necessarily follow the overdensity of dark matter, in which is used for the derivation of linear perturbation theory, it is scaled by a bias factor

$$b = \frac{\delta_g}{\delta_{dm}}, \quad (2.54)$$

where subscript g is for galaxies and dm is for dark matter. This way, one can normalize the dark matter power spectrum by measuring the $\sigma_{8,g}$ parameter and using the relation in equation 2.53 to use scales on $8\text{Mpc}/h$ as a reference for normalizing the power spectrum when the shape is known.

2.1.7 The cosmic web

On large scales the universe is isotropic and homogenous. However this does not mean that every galaxy or cosmological particle is distributed randomly. As previously discussed, matter tends to flow towards overdensities. Over time many of these overdensities will collapse into large dark matter halos. Dark matter halos contain substructures and sub halos where baryonic matter can form galaxies. The massive halos are bound together by gravity in what resembles a web-like structure known as the cosmic web. The cosmic web is made up of large filaments of baryonic and dark matter, and in between there are large voids with very low density. Figure 2.4 shows the Millenium simulation with 10^{10} dark matter particles at different redshifts $z = 0$ and $z = 18.3$ corresponding to $t = 0.21\text{Gyr}$ and $t = 13.6\text{Gyr}$ after the big bang respectively. Here one can clearly see that as time evolves the matter coalesces from a relatively even distribution of matter into a web-like structure with filaments and voids. The same pattern is also visible in the SDSS sky survey shown in figure 2.1.

An important phenomena for the matter to coalesce into this web-like structure is baryonic acoustic oscillations (BAO). Baryonic acoustic oscillations originate from the dense hot plasma that the universe consisted of after the big bang. During this time photons and baryonic matter was coupled together. While dark matter is collisionless, the baryon-photon plasma on the other hand is not and is subject to pressure forces. Gravity will cause matter to flow towards overdensities. The collisionless dark matter will clump together and form potential wells, while the baryonic matter is subject to outward pressure as it clumps together. While the photons and baryons are coupled together they move with the sound speed, which in this plasma is roughly half the speed

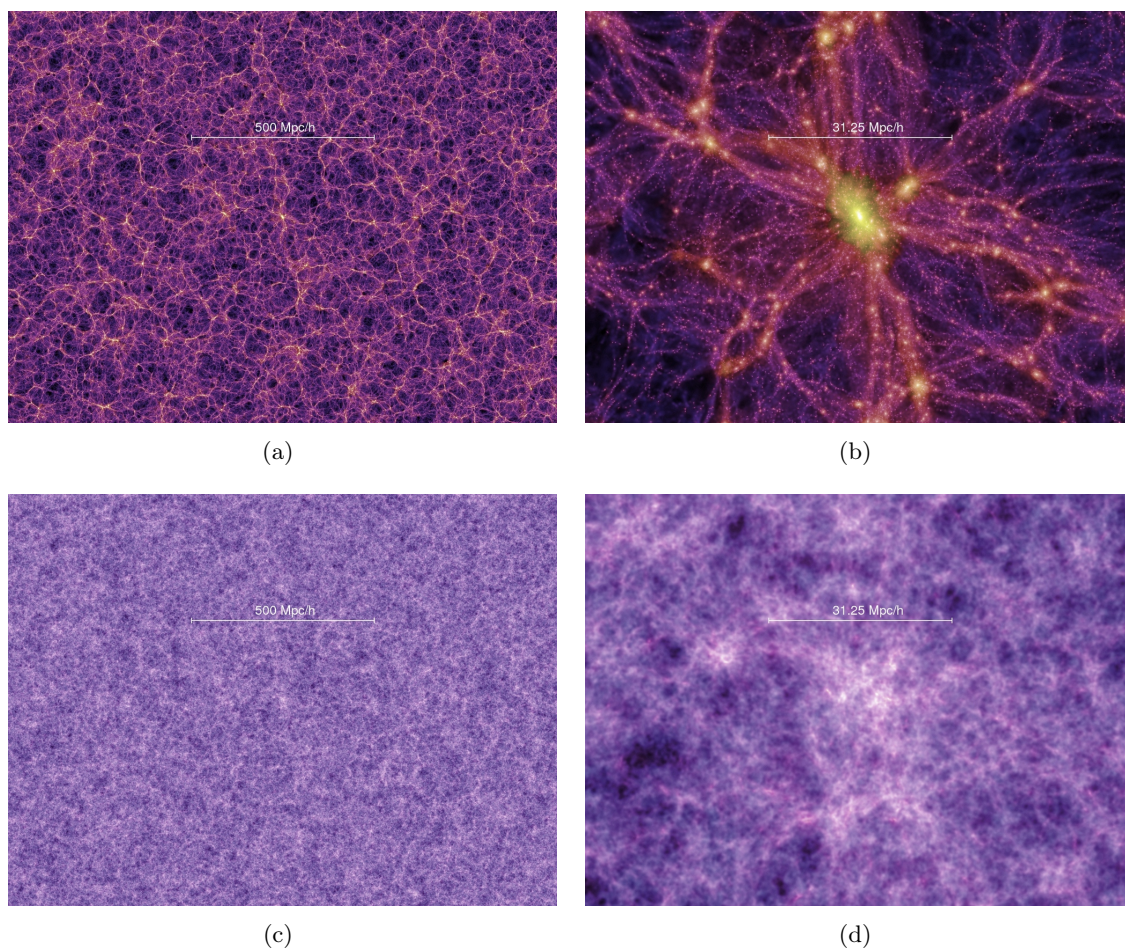


Figure 2.4: Figure showing snapshots of the Millennium simulation with 10^{10} particles for two different redshifts. Figures a) and b) is taken at $z = 0$ while figures c) and d) is taken at $z = 18.3$. Image credit: Downloaded from the website of the Millennium Simulation[2] (<https://wwwmpa.mpa-garching.mpg.de/galform/virgo/millennium/>).

of light. The outwards pressure will cause ripples of baryonic matter flowing outwards from the dark matter overdensities. At approximately $z \approx 1000$ baryons and photons decoupled. In cosmological history this period is called the epoch of recombination. Due to the expansion of the universe, the experienced interaction rate between photons and baryonic matter decreased drastically. This caused the rate at which photons scattered off of baryonic matter to drastically decrease. This allowed for the formation of neutral hydrogen. As baryonic matter was no longer bound to photons this also caused the sound speed to drop eventually causing the ripples to stop expanding. This led to shells of baryonic matter eventually attracting portions of dark matter away from the overdensities in the center of the potential wells formed before recombination. These ripples are not directly observed but can be inferred from measuring the correlation function for galaxies. The correlation function contains a bump at the characteristic length scale r_s . This shows that the baryonic acoustic oscillations are influential on the distribution of matter in the universe today.

2.1.8 Real- vs Redshift space

As mentioned earlier in section 2.1.1, due to the fact that the universe is expanding, galaxies we observe in the universe on cosmological scales move away from us and appear redshifted. In addition, galaxies also have peculiar velocities, which is their velocity when subtracting the Hubble flow. This results in an effect, that when observing galaxies as a function of redshift, their spatial distribution appears squashed. As most cosmological surveys of galaxies are done by measuring their redshift, understanding how this affects their spatial distribution is crucial. We will now denote real space coordinates by subscript r and redshift space coordinates by subscript s . By taking into account the peculiar velocity of galaxies and the Hubble flow, we can express their redshift as a sum of two terms

$$cz = H_0 d_p + v_{pec}. \quad (2.55)$$

Here the first term represents the Hubble flow and the second term is the peculiar velocity along the line of sight. For most galaxies, only the redshift can be measured and not the proper distance. Therefore, by dividing equation 2.55 by H_0 , it is useful to define the redshift distance s as

$$s = \frac{cz}{H_0} = d_p + \frac{v_{pec}}{H_0}. \quad (2.56)$$

The peculiar velocity of galaxies can also have a profound effect on how their spatial distributions appear in redshift space. Matter will cluster and move towards overdensities. This means that the velocity component of galaxies in a cluster will essentially point towards the center of the galaxy cluster. This will give an effect that when viewing a galaxy cluster, the galaxies closest to us in real space will have their peculiar velocity move away from the observer. From equation 2.55 one can see that this will give an

additional contribution in addition to the Hubble flow and the measured redshift will be higher. Galaxies on the opposite side of the cluster as seen by the observer will have their peculiar velocity point towards the observer. This will cause the peculiar velocity to contribute with a blueshift, and likewise, the measured redshift will appear smaller than if one could only measure the redshift due to the Hubble flow. This effect will cause galaxy cluster to appear squashed in redshift space. This effect will also cause the overdensity to appear much larger in the center of a collapsed object than in real space. This is illustrated in figure 2.5.

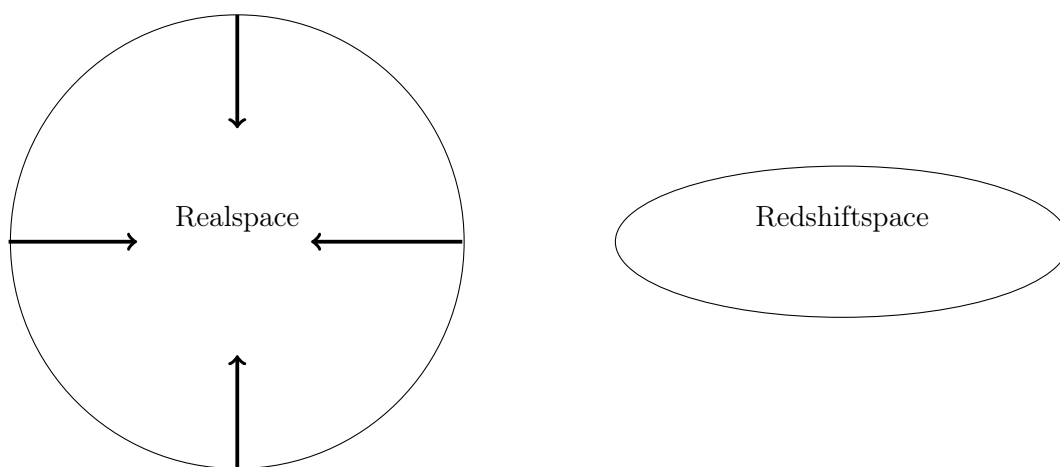


Figure 2.5: Figure illustrating a collapsing halo in realspace and redshiftspace. The vector pointers represent the direction of the peculiar velocity. In this illustration both circles are meant to contain the same number density of galaxies thereby giving the apparent overdensity appear much higher in redshiftspace than in realspace. The observer sitting at the bottom of the page with the line of sight direction being parallel to the plane spanned by the circle and the ellipse.

The effect of redshift space distortions causes the correlation function and thereby matter power spectrum to be slightly altered when observed in redshift space. In Fourier space, the overdensity receives a correction term as [24, p. 279]

$$\delta_s(\mathbf{k}) = (1 + \beta\mu^2)\delta(\mathbf{k}). \quad (2.57)$$

Here $\mu = \mathbf{z} \cdot \mathbf{k}$ is the angle between the line of sight and the Fourier wave vector. The factor β can be seen as the growth factor f introduced in equation 2.40 from section 2.1.4. However, it is scaled by the linear bias factor given in equation 2.54 as

$$\beta = f/b. \quad (2.58)$$

2.2 Cosmological N-body simulations

Analytical approaches, such as linear perturbation theory, has their limits. Especially when taking into consideration small scale interactions between galaxies. Using linear perturbation theory, gravitational interactions on small scales can not be described in enough detail to accurately compare observational data to theoretical predictions. Therefore, numerical simulations of structure formation is an important tool when studying the properties of the universe. Numerical simulations have been crucial for establishing Λ CDM as the standard model for cosmology. Only through simulations has it become possible to separate the predictions of different models when comparing models to observations [23, p. 361]. The first cosmological N-body simulation was conducted by [33], where the equations of motion for 300 particles were solved to study the formation of galaxy clusters. Due to the significant increase in computer power over the last few decades, one has been able to perform increasingly more detailed numerical simulations where one can afford to increase both the spatial and temporal dimension in the simulations.

When simulating large scale structure formation in the universe, it is mostly sufficient to take into account dark matter, as this is the primary contributor to the mass in the universe. The size of dark matter particles will be a lot smaller than the simulation volume. Therefore all the dark matter particles in the simulation are represented by point particles with mass M , where each point particle is a body of mass representing multiple dark matter particles. One also has to restrict the simulation volume of the simulation. The size of the universe is too large to simulate in detail. It may also be infinite. Therefore one has to select a simulation volume that is representative for the effects one wants to model. For simulations where one wants to examine large scale structures, the simulation volume is usually a box with size $L > 200\text{Mpc}$, as large scale structures are hardly present on scales smaller than this [23, p. 362]. Due to restricting the simulation to a small boxed slice of the universe, problems with the boundaries of the simulation volume will rise. If not treated, particles at the boundary of the simulation volume will not be affected gravitationally by as many neighbouring particles as particles in the middle of the simulations volume. To account for the fact the universe should be isotropic and homogenous on large scales one has to apply periodic boundary conditions. This means that the cube is extended periodically. A particle leaving the volume in one end will reappear in the other end. Particles will also interact gravitationally in the same manner meaning that particles on opposing sides of the simulation volume will affect each other as if the simulation volume was extended by adding an identical volume side by side.

2.2.1 Pairwise summation

The simplest and most intuitive way of performing N-body simulations is by modelling gravity as a force and applying Newton's second law to calculate the acceleration vector of the particles. For every particle, one calculates the force acting on it from all other

particles by summing over the contribution from all other particles. The computational cost of this approach scales as N^2 , where N is the number of particles in the simulation. This makes this approach not feasible when including $N \gtrsim 10^6$ particles. However interesting problems can be studied by not exceeding this number of particles. This problem can be formulated mathematically as

$$\frac{d^2 \mathbf{r}_i}{dt^2} = -G \sum_{i=1}^N \sum_{j \neq i}^N \frac{m_j}{|\mathbf{r}_j - \mathbf{r}_i|^3} (\mathbf{r}_j - \mathbf{r}_i), \quad (2.59)$$

where the acceleration of the i th particle with position \mathbf{r}_i is calculated by summing over all particles $j \neq i$ with position \mathbf{r}_j . When studying equation 2.59, one can see that if two particles come very close to each other, numerical instability will occur in the form of a division by zero, which in turn will result in an unreasonably high force calculation. To counteract this problem force softening is introduced. A way to do this is to modify the denominator of equation 2.59 by including a small factor ϵ replacing $\Delta r_{i,j} = |\mathbf{r}_j - \mathbf{r}_i|$ with $\sqrt{\Delta r_{i,j}^2 + \epsilon^2}$ [34, ch.2].

We will now use the notation where the acceleration is given as $a(t) = d^2 r(t)/dt^2$ and the velocity $v(t) = dr(t)/dt$. From equation 2.59 we now have an expression for the acceleration of each object. This gives us the coupled system ODEs necessary to solve for the position as

$$v(t) = \frac{dr(t)}{dt} \quad \text{and} \quad a(t) = \frac{dv(t)}{dt}. \quad (2.60)$$

We now discretize the time variable such that t_n is the time at a specific time point, where $t_n = t_0 + n\Delta t$, with Δt representing an incremental time step. By a Taylor expansion around t_n , one can approximate the time solution at time step $t = t_n + \Delta t$ giving

$$r(t_n + \Delta t) = r(t_n) + \frac{dr(t_n)}{dt} \Delta t + \frac{1}{2} \frac{d^2 r(t_n)}{dt^2} \Delta t^2 + \dots \quad (2.61)$$

and

$$v(t_n + \Delta t) = v(t_n) + \frac{dv(t_n)}{dt} \Delta t + \frac{1}{2} \frac{d^2 v(t_n)}{dt^2} \Delta t^2 + \dots \quad (2.62)$$

By discretizing the spatial variables in the same way as the time variable, one can write $r(t_n) = r_n$, $v(t_n) = v_n$ and $a(t_n) = a_n$ respectively for every particle. Similarly we have $r_{n+1} = r(t_n + \Delta t)$ etc. By truncating the Taylor expansion at the second order term one can then write an approximate solution to our problem as

- $a_n = - \sum_{i=1}^N \sum_{j \neq i}^N \frac{Gm_j}{(\Delta r_{i,j}^2 + \epsilon^2)^{3/2}} (\mathbf{r}_j - \mathbf{r}_i)$
- $v_{n+1} = v_n + a_n \Delta t$
- $r_{n+1} = r_n + v_n \Delta t.$

This is done for every particle i for every time step $t + \Delta t$. This is one of the simplest ways to numerically solve this system of coupled ODE's and is called the forward Euler

method. This approach however is never used in practice as the error of the solution, denoted ϵ , as can be seen from the Taylor expansion, is of order $\epsilon \sim \mathcal{O}(\Delta t^2)$. Another significant disadvantage with this method is that when solving for physical systems, it will not conserve energy. A numerical integration scheme well suited for gravitational problems where energy is conserved is the Velocity Verlet algorithm. This integration scheme is what is known as symplectic, meaning that it conserves the energy of the system [35, p. 31]. The Velocity Verlet scheme updates its velocity and position as

- $r_{n+1} = r_n + v_n \Delta t + \frac{a_n}{2} \Delta t^2$
- $v_{n+1} = v_n + \frac{\Delta t}{2}(a_{n+1} + a_n)$.

This method, and its variations, that preserve energy over long time periods makes it a preferred method of choice for astrophysical gravitational problems.

2.2.2 Particle mesh

Another way of simulating a box with a large volume of particles is to apply what is called a Particle Mesh code [36] [34]. This method is more computationally effective as the number of calculations scales proportionally to the number of particles N . However, most of the information is stored on a 3D grid covering the whole simulation volume. This will require more memory as the resolution of the grid increases.

The Poisson equation for gravity reads

$$\nabla^2 \Phi = 4\pi G \rho. \quad (2.63)$$

This equation can be solved for the gravitational potential Φ and the force field updated as $\mathbf{F} = -\nabla \Phi$. This method is reliant on assigning a density field to the simulation volume of individual particles as a representation for the density ρ in equation 2.63. One popular method for this is what is known as Cloud In Cell (CIC). One wants to calculate the density S at a given distance, given by the coordinate point (x, y, z) , from the particle. The density at a given point is given as $S(x)S(y)S(z)$, where the CIC scheme for assigning a value to S reads

$$S(x) = \frac{1}{\Delta x} \begin{cases} 1, & \text{if } |x| < \Delta x/2 \\ 0, & \text{otherwise,} \end{cases} \quad (2.64)$$

where Δx is the cell size. The density assigned to the grid is defined by a product of three weight functions $W(\mathbf{r}_p - \mathbf{r}_{ijk}) = W(x_p - x_i)W(y_p - y_j)W(z_p - z_k)$. Where subscript p denotes particle position and subscript i, j, k denotes a given grid index. A given W is expressed as

$$W(x) = \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} S(x_p - x') dx'. \quad (2.65)$$

We then have that the density assigned to a given grid point for a grid consisting of N particles is given by the relation

$$\rho_{ijk} = \sum_{p=1}^N m_p W(\mathbf{r}_p - \mathbf{r}_{ijk}), \quad (2.66)$$

where m_p is the mass of particle p . With the density assigned, one can then solve the Poisson equation for gravity given by equation 2.63. This is done by using a Fast Fourier Transform (FFT) on the whole grid giving

$$\nabla^2 \Phi = 4\pi G \rho \quad (2.67)$$

$$-k^2 \hat{\Phi} = 4\pi G \hat{\rho}, \quad (2.68)$$

where $\hat{\Phi}$ and $\hat{\rho}$ denotes the Fourier transform of the given quantities. Dividing by k^2 and taking the inverse transform one solves the equation for the gravitational potential as

$$\Phi = \text{IFFT}\left[-\frac{4\pi G \hat{\rho}}{k^2}\right]. \quad (2.69)$$

With the gravitational potential, one can calculate the force and interpolate forces to their respective particle positions in the grid. The particle mesh method follows the same algorithm as the pairwise summation method but differs in the force calculation. Summarized the force calculation of the particle mesh method consists of three steps:

- Assign density to grid.
- Solve the Poisson equation given by equation 2.63.
- Compute the force and apply to particles in the grid.

While the computation time is significantly decreased the particle mesh method lacks in terms of resolution. Increasing the resolution for a particle mesh code means increasing the number of bins in each direction. For a three dimensional grid, this will quickly put constraints on the computer memory.

2.2.3 Tree codes and Barnes-Hut simulations

The particle mesh and the pairwise summation methods are the simplest forms of N-body simulation methods. In modern codes however these methods are rarely used in exclusivity. Instead, in many modern codes, the methods are based on the approach proposed by Barnes and Hut in 1986[37]. This method divides the simulation volume into what is called a tree. An example of such a tree in 2D space is illustrated in figure 2.6. The volume in 2D is divided into four cells or eight cells in 3D. If the number of particles in each cell exceeds a certain threshold specified by the user, the given cell is then divided into four or eight more cells depending on whether the simulation is in 2D

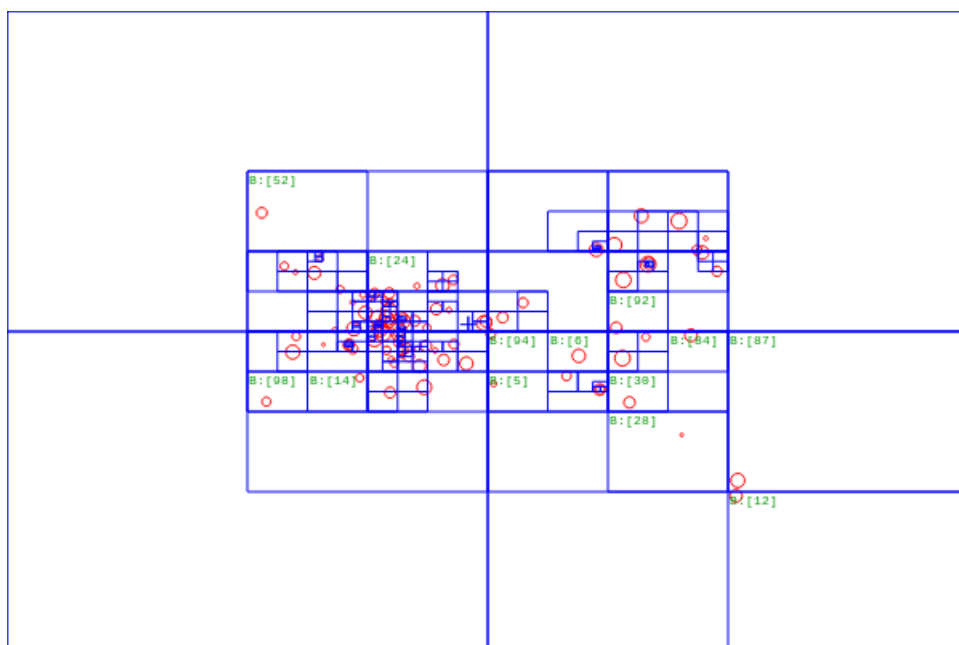


Figure 2.6: Illustration of a Barnes-Hut tree for 100 particles in 2D space. Image downloaded from https://commons.wikimedia.org/wiki/File:2D_Quad-Tree_partitioning_of_100_bodies.png. The file is licenced under the following license <https://creativecommons.org/licenses/by-sa/3.0/deed.en>.

or 3D. This dividing of cells into smaller cells continues until no cells contain a particle number higher than the set threshold. After creating the tree, the forces have to be calculated. When calculating the force on a given particle, the distances of particles in other cells becomes crucial. The basic idea is that if a certain number of particles are far away from a given particle, one can calculate the center of mass for a larger cell containing more particles. This greatly reduces the number of calculations as the force from far away particles can be approximated as a single force from one large body instead. The advantage with this approach is that, instead of the calculations scaling as $\mathcal{O}(N^2)$ as for the regular pairwise summation, it scales as $\mathcal{O}(N \log N)$ instead. This approach briefly described here is fundamental for the approach for many high-level cosmological N-body simulation codes.

2.2.4 Initial conditions

When setting the initial conditions for cosmological simulations, one starts at a high redshift. The initial distribution of the particles resembles a gaussian random field with the theoretical matter power spectrum $P(k, z)$ for the given cosmological model. Choice of a time step is also important. One has to find a compromise between long computation time and accuracy between closely interacting particles.

2.3 Statistics

For the analysis of the data, statistical methods are utilized. This section gives a brief overview of the fundamental methods used for basic statistical analysis. This section is largely based on [38] and [39].

2.3.1 Bayes theorem

When studying and comparing our observations and simulations with theory one has to quantify the probability of the models parameter space given the data used. I.e, what parameters will make our model compare best with our observations or simulations. Consider the probability $P(A, B)$, which is the probability for two events A and B occurring. If the two events A and B are independent of each other, the probability is simply given as $P(A, B) = P(A)P(B)$. However, this is not the case if the events A and B are dependent on each other. Take for instance the probability of drawing two knights in a row from a deck of cards. The first draw alters the number of cards in the deck and will therefore affect the probability of the second draw. We write this as $P(A, B) = P(A)P(B|A)$. Here $P(B|A)$ denotes the probability of B given the fact that event A has occurred. We can likewise write $P(A, B) = P(B)P(A|B)$. Now let's consider the event A being the parameters of our model θ , and event B being our data represented by d . We can then write $P(\theta, d) = P(\theta)P(d|\theta) = P(d)P(\theta|d)$. Rearranging terms we will get what is known as Bayes theorem[38]

$$P(\theta|d) = \frac{P(d|\theta)P(\theta)}{P(d)}. \quad (2.70)$$

$P(\theta|d)$ is what is known as the posterior and quantifies an important question one would ask when studying cosmology or physics in general. What is the probability of the parameters of the model being correct given the data one is studying? Likewise $P(d|\theta)$ is the probability of the data being true given the parameters one has tested. This is what is known as the likelihood and is usually denoted as $\mathcal{L}(\theta)$. $P(\theta)$ is what is known as the prior. This is used to quantify what we already know about our parameter space. $P(d)$ is what is known as the evidence and acts as a normalization factor.

In most practical examples our model contains multiple parameters we want to fit. Let's say we have a model dependent on the parameters θ_0 and θ_1 . We will then get the joint probability distribution $P(\theta_0, \theta_1)$. If we want to know the probability of θ_0 independently of θ_1 , we can get what is known as the marginal distribution given by

$$P(\theta_0) = \int_{-\infty}^{\infty} P(\theta_0, \theta_1) d\theta_1 \quad (2.71)$$

2.3.2 Parameter estimation

When comparing models with data, one has to find an expression to quantify the agreement between the model and the data. An important statistical tool is what is called the χ^2 test. Given two functions y_i^{model} and y_i^{data} , representing the model and data respectively, where i is a given data point, the χ^2 becomes

$$\chi^2 = \sum_{ij} (y_i^{\text{model}} - y_i^{\text{data}}) C_{ij}^{-1} (y_j^{\text{model}} - y_j^{\text{data}}). \quad (2.72)$$

In this expression, C_{ij} is the covariance matrix if the data are correlated. If the data is uncorrelated, the χ^2 takes a simpler form

$$\chi^2 = \sum_i \frac{(y_i^{\text{model}} - y_i^{\text{data}})^2}{\sigma_i^2}, \quad (2.73)$$

where σ_i is the error estimate for the given data point i . This is a useful tool for testing the scatter of our data. The likelihood function $\mathcal{L}(\theta) = P(d|\theta)$ describes the probability of our dataset d being true given our parameters. By holding our dataset constant one can maximize the likelihood function. In other words, one can try to find the most probable value for θ . If one assumes the data is given by a gaussian distribution, the likelihood function is defined by a multivariate gaussian

$$\mathcal{L}(\theta) = \frac{1}{\sqrt{(2\pi)^n |C|}} \exp\left[-\frac{1}{2} \sum_{ij} (y_i^{\text{model}} - y_i^{\text{data}}) C_{ij}^{-1} (y_j^{\text{model}} - y_j^{\text{data}})\right], \quad (2.74)$$

where n is the number of parameters. From this, we can see that the likelihood and the χ^2 is related as $\mathcal{L} \propto \exp[-\frac{1}{2}\chi^2]$. From this, it is also evident that maximizing the likelihood is the same as minimizing χ^2 . Using Bayes theorem given by equation 2.70 one can see that we now have an expression for the likelihood. The evidence $p(d)$, which

acts as a normalization factor is not important as we want the relative probabilities of our parameters and therefore this term can be ignored [40]. If one chooses a flat prior, one where $p(\theta)$ is constant, one simply gets by addressing Bayes theorem that

$$P(\theta|d) \propto \mathcal{L}(\theta). \quad (2.75)$$

2.3.3 Sampling and Metropolis-Hastings MCMC

When dealing with distributions $P(\theta, \mathbf{d})$, where there is a large amount of parameters $\theta = (\theta_0, \theta_1, \dots, \theta_n)$, a simple gridding algorithm can prove to be both time and memory consuming. If we also take into account the fact that one may deal with gaussian distributions, there are large areas of parameter space where the probability of particular configurations of parameters being true are slim to none. Therefore, by gridding, one spends a significant amount of time exploring a relatively uninteresting part of parameter space. Therefore, it is useful to introduce the concept of sampling and the Metropolis-Hastings Markov-Chain Monte Carlo method.

The Monte Carlo principle states that if you take independent and identically distributed samples θ from an arbitrary function $f(\theta)$, as the number of samples increases, the distribution of samples will converge towards the function $f(\theta)$. This is expressed mathematically as

$$f_N(\theta) = \frac{1}{N} \sum_{i=1}^N \delta(\theta) \quad (2.76)$$

$$\rightarrow \lim_{n \rightarrow \infty} f_N(\theta) = f(\theta) \quad (2.77)$$

There is however a problem in that the function we want to calculate $P(\theta|d)$ is a distribution that is unknown to us. We know how to calculate it, but we cannot draw from it. Therefore one introduces a proposal distribution $Q(\theta)$. This is a known function that one can easily evaluate and pick samples from. The proposal distribution can be any function, but it is usually a Gaussian or any other function enclosing the parameter space of the target function. The target function is the function one wants to represent, but it may be computationally ineffective to use a brute force grid approach if the model contains a multitude of parameters. In our case, the target distribution is the likelihood or the posterior distribution.

One of the most popular Markov chain Monte Carlo algorithms is the Metropolis-Hastings algorithm. The basic idea behind the Metropolis-Hastings algorithm is to draw a point θ^* from the proposal distribution $Q(\theta)$ based on the previous draw, and compare the draw with an acceptance criterion. The probability for the acceptance criterion is usually drawn from a uniform distribution. The acceptance criterion in Metropolis-Hastings is given by

$$\alpha = \min\left[1, \frac{P(\theta^*)Q(\theta^*|\theta)}{P(\theta)Q(\theta|\theta^*)}\right], \quad (2.78)$$

where $Q(\theta^*|\theta)$ represents the proposal distribution for a new pick of parameters θ^* based on previous parameters. $P(\theta)$ is the target density distribution. After drawing an initial value θ_0 from the proposal distribution Q , the Metropolis-Hastings algorithm is a repeatable sequence of the following

- Sample θ^* from $Q(\theta^*|\theta)$

- Sample a number u from the normal distribution

- if $u < \alpha$, where α is given by equation 2.78:
accept move and sample θ^*

- else:
reject the move and sample the previous point θ instead.

- Repeat until you have a sufficient amount of samples to accurately represent the target distribution.

The idea with the acceptance criterion is that the algorithm will eventually explore all of the parameter space where the target distribution has its highest values. If the algorithm just accepted the next step without comparing it to a random pick from the normal distribution, the algorithm would just converge to the closest peak. By comparing the acceptance criterion to a pick from the normal distribution, one would also in some cases accept picks with a less probable state. This makes the algorithm map the whole distribution, but it will quickly drift away from spending too much time exploring parameter space where the target distribution is negligible.

Chapter 3

Codes utilized - Identifying voids and filaments from a particle distribution

3.1 DisPerSE-"Discrete Persistent Structures Extractor"

For the analysis conducted on filaments, the identification of the filamentary structure of the cosmic web in data sets is crucial. To achieve this, the code DisPerSE[41][42] is utilized. The code is publicly available at <http://www2.iap.fr/users/sousbie/web/html/indexd41d.html?>. DisPerSE is a code that was developed to identify topological features of the cosmic web in cosmological particle distributions. DisPerSE utilizes discrete Morse Theory to identify the topological structures of the cosmic web in data sets. In this section, I will give a brief overview of how DisPerSE works to identify the filamentary structure from a particle distribution. For a thorough review I recommend the paper [41], in which the following chapter is based and much notation is borrowed, describing the theory behind DisPerSE and the references therein. Another good summary of the beforementioned paper is found in [43], from which the following chapter also has drawn information.

3.1.1 Morse Theory

To identify the features of the cosmic web from the density field of our particle distribution, DisPerSE relies on Morse Theory [44]. Morse theory is a mathematical framework used to relate the geometrical and topological properties of a function. In our case, the topological features are quantities such as the number of galaxy clusters or dark matter halos, while the geometric features are quantities related to the precise position of the galaxy cluster or dark matter halos and the local density. In this thesis, Morse theory is utilized to study the relation between the density field derived from a dark matter halo catalogue and features such as filaments, voids and walls. This is done by studying the gradient of a smooth function f . By studying the gradient of the function f , one can

analyze the topological and geometrical features of the manifold. In general, to apply Morse theory, one needs a smooth scalar function f that is twice differentiable. In our case, this scalar function is the density field of our simulation volume. In Morse theory, the gradient of the function f is defined as $\nabla_x f(\mathbf{x}) = df(\mathbf{x})/d\mathbf{x}$. This function specifies the direction of the steepest ascent for our scalar density function. The points where $\nabla_x f = 0$, are called critical points. These critical points can be classified by studying the Hessian matrix of f given as

$$\mathcal{H}_f(\mathbf{x}) = d^2 f(\mathbf{x})/dx_i dx_j. \quad (3.1)$$

The critical points are classified by the eigenvalues of $\mathcal{H}_f(\mathbf{x})$. A critical point is of order k when the Hessian matrix has exactly k negative eigenvalues. This way one can classify the critical points as maxima with $k = 2$, saddle point with $k = 1$ and minima with $k = 0$. This analysis also implies that for Morse theory to be applicable, the function f has to satisfy the condition $\mathcal{H}_f(\mathbf{x}) \neq 0$, where $\nabla_x f = 0$. From this, it follows that any function that satisfies this condition can be classified as a Morse function.

It is previously stated that the gradient points in a preferred direction for any non-critical point. From the gradient, one can therefore define integral lines between the critical points. An integral line is a parametrised curve $L(t) \in \mathbb{R}^d$ that satisfies

$$\frac{dL(t)}{dt} = \nabla_x f(\mathbf{x}). \quad (3.2)$$

These integral lines will always have critical points as their origin and destination. An important property of these integral lines is that they cover all of \mathbb{R}^d , but they never intersect each other. They can however share the same origin or destination. With these integral lines, one can define ascending and descending n -manifolds. If one considers a critical point P of order k on the Morse function living in \mathbb{R}^d , an ascending manifold of $d - k$ dimensions is constructed by the set of points reached by all integral lines with origin at the critical point P . A descending manifold is a k -dimensional region of space defined by all integral lines with their destination at point P . An ascending 3-manifold associated with minimas trace underdense regions represented by voids. Ascending 2-manifolds trace walls and the filaments themselves are traced by ascending 1-manifolds. This leads to the definition of the Morse complex. The Morse complex is simply the set of all the ascending or descending manifolds. The Morse complex establishes the link between the geometrical properties such as the location of critical points and the paths of the gradients with the topological feature such as how the critical points are connected and how many there are of each type.

3.1.2 Discrete Morse theory

The formality currently introduced deals with smooth continuous functions. This however is rarely applicable to measured data or simulations. For example, the density field of our simulation volume is derived from a discrete set of point particles. Therefore

discrete Morse theory[45] is applied in order to utilize Morse theory on discrete data. Instead of working on smooth functions, discrete Morse theory is applied to what is called a simplicial complex. A simplicial complex is a space made up of simplices. A simplex is a generalization of a triangle to an arbitrary number of dimensions. For example a k -dimensional simplex, referred to as a k -simplex, is represented as

- a 0-simplex is a point
- a 1-simplex is a line segment
- a 2-simplex is a triangle
- a 3-simplex is a tetrahedron/pyramid etc.

A k -simplex will be denoted as σ_k defined by the set of points $\sigma_k = \{p_0, \dots, p_k\}$. Similarly, one can define the face of a simplex to be a subset of the original k -simplex. We can define the face as an l -simplex $\gamma_l = \{p_0, \dots, p_l\}$ with $l \leq k$. If γ_l is a face of σ_k , then σ_k is called a coface of γ_l . When k and l only differs by one, a face is called a facet and a coface is called a cofacet. These simplices make up what is called a simplicial complex in which a discrete Morse function is defined and discrete Morse theory is applied. A simplicial complex K is made up of a finite set of simplices such that $\sigma_k \in K$. With this one can define a discrete Morse function f as a function that maps a real value $f(\sigma_k)$ for each simplex in the simplicial complex K . As for the Morse function defined in regular Morse theory, the discrete counterpart also has to be differentiable and the gradient can only flow towards one preferred direction locally. For this to occur and a Morse function to be defined on the simplicial complex it has to satisfy the following criteria

- there exists at most one facet ξ_{k-1} of σ_k such that $f(\sigma_k) \leq f(\xi_{k-1})$, and
- there exists at most one cofacet χ_{k+1} of σ_k such that $f(\sigma_k) \geq f(\chi_{k+1})$.

These criteria states that locally a simplex has a lower value than its facet and a higher value than its cofacet. On a discrete Morse function, one defines a gradient field. This is done by comparing a k -simplex with its facets or cofacets. One can couple a k -simplex with its facets or cofacet and identify a preferential flow direction through gradient pairs

- If a simplex σ_k has exactly one lower valued cofacet χ_{k+1} , then $\{\sigma_k, \chi_{k+1}\}$ defines a gradient pair
- If a simplex σ_k has exactly one higher valued facet ξ_{k-1} , then $\{\xi_{k-1}, \sigma_k\}$ defines a gradient pair.

Gradient pairs define gradient arrows and define the preferred flow direction in the discrete vector field. As for the regular Morse theory, the discrete counterpart also requires an analogy to a critical point. These are called critical k -simplices. A k -simplex σ_k is critical if the following criteria are satisfied

- there exists no facet ξ_{k-1} of σ_k such that $f(\sigma_k) \leq f(\xi_{k-1})$, and
- there exists no cofacet χ_{k+1} of σ_k such that $f(\sigma_k) \geq f(\chi_{k+1})$.

In regular Morse theory, the integral lines were crucial in representing the manifolds in which the Morse complex is defined. Their discrete counterpart is what is called a V-path. If one considers k -simplexes α_k^i and $(k + 1)$ simplexes β_{k+1}^j , where α_k^{i+1} is a facet of β_{k+1}^i , a V-path is a strictly decreasing an alternating sequence such that

$$\alpha_k^0, \beta_{k+1}^0, \alpha_k^1, \beta_{k+1}^1, \dots, \alpha_k^n, \beta_{k+1}^n. \tag{3.3}$$

Each pair $\{\alpha_k^i, \beta_{k+1}^i\}$ forms a gradient pair. A V-path is a set of all continuous gradient pairs through the simplicial complex K . With V-paths one can define discrete n -manifolds that make up the discrete Morse complex. A discrete n -manifold is the set of k -simplices that belongs to a V-path with either origin or destination at the critical simplex σ_k . The n -manifolds are ascending if their origin is σ_k . Likewise, the n -manifolds are descending if their destination is σ_k . These manifolds make up what is called the discrete Morse complex.

3.1.3 Delaunay Tessellation Field Estimator

In order to compute a density field from a set of point particles, in which one can compute the discrete Morse complex, DisPerSE utilizes Delaunay-tessellation[46]. The Delaunay Tessellation Field Estimator (DTFE) uses Delaunay triangulation. The Delaunay triangulation divides the volume containing the particles into a set of triangles in 2D or tetrahedrons 3D. The volume is divided such that no particle resides inside the circum-circle of the assigned triangles. This triangulation captures the density of the field in such a way that if the triangles are small, there are many particles close to each other and the density is high. Likewise, larger triangles are made up of a point distribution with a larger distance between the particles and the density is lower. After assigning a density to each triangle, one can use linear interpolation, assuming the density changes linearly between each bin, to assign a density to all bins in the assigned computational grid.

3.1.4 Topological Persistence

When identifying the structures of the cosmic web from a density field derived from a point distribution, one may encounter noise. Structures not supposed to be significant may be interpreted as so and provide noise for later analysis. To deal with noise DisPerSE utilizes persistence theory [47]. Persistence theory studies the evolution of what is called excursion sets. An excursion set, or a sub-level set, is given as

$$(x_1, \dots, x_n) | \rho(x_1, \dots, x_n) \geq \rho_0. \tag{3.4}$$

This is the set of points (x_1, \dots, x_n) where $\rho(x_1, \dots, x_n)$ is larger than some threshold value ρ_0 . Persistence theory measures the so-called lifetime of the excursion set. This is a

measure of the absolute difference in value between a pair of two critical points. By cycling through values for ρ_0 one may measure the lifetime of the topological features. The summary of DisPerSE presented by [43] provides an analogy for this using an example with a function represented as a landscape submerged under water. I will borrow this analogy here. One can imagine a function of one variable $\rho(x)$ with peaks and troughs as mountains submerged under a water level set by ρ_0 . This is illustrated in figure 3.1. When the whole function is submerged under water the excursion set is empty. As the water level ρ_0 decrease the first highest peak of the function will eventually become visible and a new component of the excursion set will start to grow. Eventually the second peak will appear and a new independent component will start to grow. These two components are independent of each other because, as seen from above the surface of the water, they are still separated. These two components will grow until the through separating the two peaks will become visible. If the persistence of the pair is lower than a given threshold the two independent components of the excursion set will now merge together and form a persistence pair. When two components merge one is destroyed. In this case the lowest maximum is destroyed by the minimum as is illustrated in figure 3.1.

DisPerSE cancels noise in the dataset by computing the persistence pairs in the discrete Morse complex. Given a persistence pair composed of two simplicial complexes $q_k = \{\sigma_k, \sigma_{k+1}\}$ on the discrete Morse function f , a persistence ratio r is given as

$$r(q_k) = \frac{f(\sigma_{k+1})}{f(\sigma_k)}. \tag{3.5}$$

This ratio is used for determining the statistical significance of persistence pairs. The significance given in "number of sigmas" reads

$$S(q_k) = \text{Erf}^{-1} \left[\frac{P_k(r(q_k)) + 1}{2} \right], \tag{3.6}$$

where Erf is the error function and $P_k(r(q_k))$ is the cumulative probability that a persistence pair with persistence ratio $r \geq r_0$ exists in the dataset. The code takes a σ input for specifying a threshold for cutting persistence pairs. The argument specifies the persistence threshold in number of sigmas and any persistence pair with a probability lower than the given sigma threshold will be cancelled.

3.2 REVOLVER-REal-space VOId Locations from surVEy Reconstruction

For much of the void analysis conducted in this thesis the code REVOLVER-REal-space VOId Locations from surVEy Reconstruction is utilized. This code is publicly available at <https://github.com/seshnadathur/Revolver>. In the scope of this project the code provides an important utility for the analysis. It provides code for finding voids in a galaxy or halo catalogue. It also has the ability to reconstruct real space positions for redshift space positions.

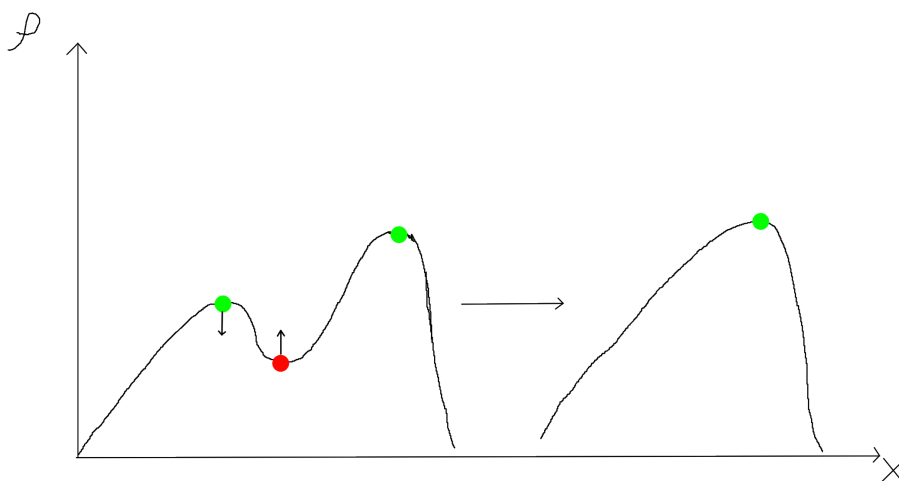


Figure 3.1: Example on how DisPerSE uses persistence pairs to simplify topological features. On the left is the unmodified function where the two leftmost critical points forms a persistence pair. The value of the minima is increased while the value of the maxima is increased. This is done until the value of the minima surpasses that of the maximum and thus the persistence pair is cancelled. This effectively smoothes out features of the function.

3.2.1 ZOBOV-ZOnes BOrdering on Voidness

The void finding method provided by REVOLVER utilized in this project is based on the ZOBOV-ZOnes BOrdering on Voidness algorithm. The ZOBOV algorithm was first presented in [48], in which this subsection is based on. In order to assign a density to the grid from the point particle distribution ZOBOV takes use of Voronoi-tesselation. Voronoi tessellation divides the space into individual cells where each cell belongs to a single particle from the point distribution. Each cell is divided so that a cell around particle i is made up of all points in space closer to that particle than any other particle. One can imagine individual bubbles at each particle in the distribution expanding at an equal rate. Where these bubbles collide defines the edges of each Voronoi cell. The density of each cell is then $\rho_i = 1/V_i$, where index subscript i represents the density and volume at each cell respectively.

After density has been assigned to the grid through Voronoi tessellation the particles are divided into zones. The zones are divided around each density minimum. The center of each zone is a Voronoi cell with a density lower than all of its neighbors. These centers are found by tracing each particles neighbors until one reaches such a particle with density lower than all its neighbors. A zone is defined as the set of particles in which the density flows downwards towards the minimum. Due to noise present in the dataset, multiple zones may make up one void. Therefore some of them may need to be joined together. The way zones are joined is considering, for a 2D density field and imagining the density function as a scalar function representing a landscape, individual zones and imagine filling them with water. For each zone the water level is set to the density at the core of each zone. Gradually raising the water level other zones will get filled. The process stops when a zone with a lower core density gets filled, or, if the current zone is the one with the lowest density, when the whole field is filled with water. The water level in which the water flows into a deeper zone is recorded as $\rho(z)$, where z denotes a given zone. This will lead to a very large void (the zone with the lowest density at the center), and many sub-voids. Now one has to define the edges of the voids to determine which zones should be counted as voids.

One method is to determine the statistical significance of voids. The probability of wether a zone should be considered as real is determined by the ratio of the density contrast of a zone with the minimum density of the zone. By comparing this to a Poisson particle distribution one can get the estimate for a likelihood function P which determines the probability that a void with a given density contrast could arise from Poisson noise. One can then choose to cut all voids exceeding a given significance level from the dataset.

3.2.2 Redshift space reconstruction

Due to all observational data being measured in redshift space, in order to study the real space positions of observed quantities one has to apply reconstruction. REVOLVER contains a numerical implementation of the algorithm described in [49] and [50], in which

describes the details of how REVOLVER reconstructs real space positions from redshift space. I will give a short recap of this method here.

The algorithm of reconstruction is derived in the framework of Lagrangian perturbation theory. In this framework the Eulerian position $\mathbf{x}(t)$ is given as

$$\mathbf{x}(t) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q}, t), \quad (3.7)$$

where \mathbf{q} is the initial Lagrangian position and $\mathbf{\Psi}(t)$ is the displacement field. The Eulerian position separates itself from the Lagrangian position in the sense that the Eulerian position is a specific point in space while the Lagrangian position is the position of a parcel where the observer follows that individual parcel with the velocity field through space and time. To first order, the displacement field $\mathbf{\Psi}$ can be modelled together with the overdensity as $\nabla \mathbf{\Psi}^{(1)}(\mathbf{q}, t) = -\delta(\mathbf{x}, \mathbf{t})$, where $\delta(\mathbf{x}, t)$ is the usual overdensity parameter given in equation 2.33. In order to extract the displacement field from $\mathbf{\Psi}(\mathbf{q}, t)$ from a redshifted density distribution, one has to solve [51]

$$\nabla \cdot \mathbf{\Psi} + \beta \nabla \cdot (\mathbf{\Psi} \cdot \hat{r}) \hat{r} = -\frac{\delta_g}{b}, \quad (3.8)$$

where β and the bias b is defined in equations 2.58 and 2.54. The vector \hat{r} is the unit vector in the radial direction. Using a Helmholtz decomposition, one can express $(\mathbf{\Psi} \cdot \hat{r}) \hat{r}$ as a divergence and a curl component, giving $(\mathbf{\Psi} \cdot \hat{r}) \hat{r} = \nabla A + \nabla \times \mathbf{B}$. By approximating $(\mathbf{\Psi} \cdot \hat{r}) \hat{r}$ to be irrotational one can neglect the curl component. Using this and rewriting $\mathbf{\Psi}$ as the gradient of a potential $\mathbf{\Psi} = \nabla \phi$, one can rewrite equation 3.8 as

$$\nabla^2(\phi + \beta A) = -\frac{\delta_g}{b}. \quad (3.9)$$

This equation can be transformed to Fourier space allowing us to write

$$\hat{\phi} + \beta \hat{A} = \frac{1}{k^2} \frac{\hat{\delta}_g}{b}. \quad (3.10)$$

In this equation, $\hat{\phi}$, \hat{A} and $\hat{\delta}_g$ denotes the Fourier transform of the quantities. Remembering that $\mathbf{\Psi} = \nabla \phi$, one can multiply by $i\mathbf{k}$ and get

$$\hat{\Psi} + i\mathbf{k} \hat{A} = \frac{1}{k^2} i\mathbf{k} \frac{\hat{\delta}_g}{b}. \quad (3.11)$$

The previous assumption that $(\mathbf{\Psi} \cdot \hat{r}) \hat{r}$ is irrotational allows us to write $(\mathbf{\Psi} \cdot \hat{r}) \hat{r} \approx \nabla A$. By taking the inverse Fourier transform of equation 3.11, one can approximate equation 3.8 as

$$\mathbf{\Psi} + (\mathbf{\Psi} \cdot \hat{r}) \hat{r} = -\nabla \nabla^{-2} \frac{\delta_g}{b}. \quad (3.12)$$

Using fast Fourier transforms this equation can be solved as [50]

$$\mathbf{\Psi} = \text{IFFT} \left[-\frac{i\mathbf{k}\delta(\mathbf{k})}{k^2 b} \right] - \frac{f}{1+f} \left[\text{IFFT} \left[-\frac{i\mathbf{k}\delta(\mathbf{k})}{k^2 b} \right] \cdot \hat{r} \right] \hat{r}. \quad (3.13)$$

Due to redshift space distortions only appearing along the line of sight direction, the component of displacement field responsible for redshift space distortions is given by $\Psi_{RSD} = -f(\Psi \cdot \hat{r})\hat{r}$.

REVOLVER solves equation 3.13 using fast Fourier transforms. Using a Voronoi tessellation for estimating the density field $\delta(\mathbf{x})$ of the simulation volume, this field is then transformed to Fourier space as $\delta(\mathbf{k})$. By then dividing by bk^2 and multiplying by $i\mathbf{k}$, one will retrieve the expression inside the braces in the equation above. The inverse Fourier transform is then taken.

Chapter 4

Dataset - Multidark Simulation

For this analysis, I will use The Big MultiDark Planck simulation (BigMDPL)[22]. The data from this simulation can be found at the cosmosim website ¹. This is a dark matter particle simulation following the evolution of 3840^3 particles starting from redshift $z = 100$. A full list of parameters, as listed on the webpage, of the dataset can be seen in table 4.1. This simulation has been carried out utilizing the L-GADGET-2 code. This code is a modification of the treeSPH[52] code GADGET-2[53] in order to account for a very large amount of particles. The simulation data utilized in this analysis is taken at $z = 1$. After simulating the evolution of the dark matter particle field from $z = 100$ to $z = 1$ a halo finder is used to identify dark matter halos in the particle field.

4.1 Halo finding using Rockstar

In order to identify dark matter halos from the dark matter particle field, the Rockstar halo finder[54] is utilized. In short, this halo finder is comprised of six steps which are summarized in the beforementioned paper. I will recap these steps here.

1. The first step is to group particles in the simulation volume using a friends of friends (FOF) algorithm. In short a FOF algorithm groups particles by using a linking length l and for every particle it groups it together with all other particles within the linking length l .
2. For each group the particles are normalized by the group position and velocity dispersions.
3. In phase-space a linking length is chosen so that 70% of the groups particles are linked together.
4. This process of steps 2. and 3. repeats itself

¹The BigMDPL dataset can be found at <https://www.cosmosim.org/cms/simulations/bigmdpl/>

Box size	2.5 Gpc/h
Number of particles	3840^3
DM particle mass	$2.359 \cdot 10^{10} M_{\odot}/h$
Force resolution	30kpc/h (high redshift) - 10kpc/h (low redshift)
Initial redshift	$z = 100$
Hubble parameter	$h = 0.677$
Ω_{Λ}	0.692885
Ω_m	0.307115
Ω_b	0.048206
Spectral index	$n = 0.96$
$\sigma_8(z = 0)$	0.8228

Table 4.1: Parameters of the BigMDPL simulation extracted from the website <https://www.cosmosim.org/cms/simulations/bigmdpl/>.

Property	MultiDark 1	MultiDark 2	MultiDark 3	MultiDark 4
Mass cut [M_{\odot}/h]	$10^{13.5}$	10^{13}	$10^{12.75}$	$10^{12.5}$
Nr of halos	534559	3460116	7523817	15365856
Bias (b)	3.85	2.77	2.40	2.11
Number density [$(h/\text{Mpc})^3$]	$0.34 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$	$4.8 \cdot 10^{-4}$	$9.8 \cdot 10^{-4}$

Table 4.2: Relevant properties for the analysis conducted in this thesis for each individual dataset after applying mass cuts to the halo catalogue.

5. Once all subgroups are calculated seed halos are placed at the lowest substructure levels. Particles are then assigned to the closest seed halo in phase space.
6. Once particles are assigned to a seed halo unbound particles are removed and the properties of the halo itself is calculated (i.e mass, velocity, position etc).

4.2 Applying mass cuts to the dataset.

After the halo finder algorithm has been applied to the dataset we are left with a catalogue of dark matter halos with x , y and z positions, v_x , v_y and v_z velocity components and the mass of each individual halo. This dataset will be pruned into four different catalogues by applying different mass cuts. Future and present observational missions will be able to locate a different number galaxies in their surveys. It is therefore important to test the performance of the methods on datasets of different sizes to see how this affects the final results. The mass cuts are done by removing all halos below a certain mass threshold from the catalogue. These will be referred to later in the thesis as MultiDark 1 (MD1), MultiDark 2 (MD2), MultiDark 3 (MD3) and MultiDark 4 (MD4) respectively. Their individual properties are listed in table 4.2.

Chapter 5

Method

5.1 Radial velocity components in linear theory

One of the goals of this project is to test the viability of using the assumptions from linear theory. Linear theory can be applied to study whether it is a good approximation for the behaviour of the matter distribution around voids and filaments. The radial velocity component of halos in filaments and voids is an analytical tool that is useful when comparing with data from simulations. The velocity field around these objects can also be an integral part when modelling the statistical properties of these objects. Testing whether linear theory can be used to approximately model the physical behavior of halos and galaxies observed around these objects is therefore crucial.

5.1.1 Radial velocity component for filaments

When calculating the radial velocity component of filaments, the average filament is approximated to be a cylindrical object. The radial component is the component of the velocity vector \mathbf{v} that points perpendicular to the longitudinal axis L . This is also called the spine of the filament. A figure of this configuration is illustrated in figure 5.1. For this analysis, we will assume that the other velocity components are negligible and that the radial component is the significant component. The continuity equation

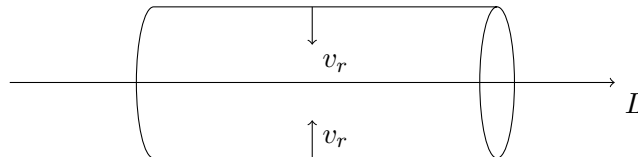


Figure 5.1: Illustration of a cylindrical filament with longitudinal axis L and radial velocity component v_r .

in linear theory can be written as

$$\frac{d\delta}{dt} + \frac{1}{a}\nabla \cdot \mathbf{v} = 0. \quad (5.1)$$

For the continuity equation in linear theory one can write $d\delta/dt = Hf\delta$ [23, p. 347], where f is the growth factor defined in equation 2.40. Substituting this into equation 5.1 one will get

$$\nabla \cdot \mathbf{v} = -Hfa\delta. \quad (5.2)$$

Assuming cylindrical coordinates we can replace the divergence with

$$\nabla \cdot \mathbf{v} = \frac{1}{r} \frac{\partial}{\partial r}(rv_r) + \frac{1}{r} \frac{\partial v_\phi}{\partial \phi} + \frac{\partial v_L}{\partial L}. \quad (5.3)$$

Assuming that the radial component v_r gives the significant contribution and the two other are negligible, by inserting into equation 5.2 one will get

$$\frac{1}{r} \frac{\partial}{\partial r}(rv_r) = -Hfa\delta. \quad (5.4)$$

Multiplying both sides by r and integrating one will be left with

$$v_r = -\frac{1}{r}Hfa \int_0^r \delta(x)xdx. \quad (5.5)$$

By defining the average mass density contrast for filaments $\Delta_f(r)$ as

$$\Delta_f(r) = \frac{2}{r^2} \int_0^r \delta(x)xdx, \quad (5.6)$$

one can write the radial velocity component for a halo in a filament as

$$v_r(r) = -\frac{1}{2}Hfar\Delta_f(r). \quad (5.7)$$

5.1.2 Radial velocity component for voids

The radial velocity component for voids is similar to the one for filaments except that for voids one assumes a spherical coordinate system. This assumption is based on the notion that the average void should have a spherical shape with, on average, no preferred direction for overdensities on the rim and thereby having a spherical density profile. Since objects cluster towards larger densities, moving straight away from the underdensity in the center of the void, one can argue that the radial velocity should be the dominating component. This is illustrated in figure 5.2. The derivation for the radial velocity component is the same for voids as for filaments up until equation 5.2. The difference for voids is that the divergence now has to be calculated in spherical coordinates given by

$$\nabla \cdot \mathbf{v} = \frac{1}{r^2} \frac{\partial}{\partial r}(r^2v_r) + \frac{1}{r\sin\theta} \frac{\partial}{\partial \theta}(\sin\theta v_\theta) + \frac{1}{r\sin\theta} \frac{\partial}{\partial \phi}(\phi v_\phi). \quad (5.8)$$

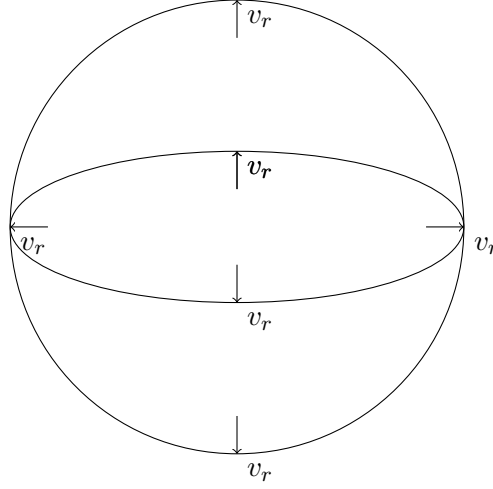


Figure 5.2: Illustration of a spherical void with radial velocity pointing outwards from the center of the void.

As for filaments, only the radial component is assumed to be non-negligible giving

$$\frac{1}{r^2} \frac{\partial(r^2 v_r)}{\partial r} = -H f a \delta. \quad (5.9)$$

Multiplying by r^2 and integrating one will get

$$v_r = -\frac{1}{r^2} H f a \int_0^r x^2 \delta(x) dx. \quad (5.10)$$

Similar to what was done for filaments, the average mass density contrast for voids is defined as

$$\Delta_v(r) = \frac{3}{r^3} \int_0^r x^2 \delta(x) dx, \quad (5.11)$$

which will give an expression for the radial velocity for voids

$$v_r(r) = -\frac{1}{3} H f a r \Delta_v(r). \quad (5.12)$$

5.1.3 Minor correction term to velocity around voids.

The equation for the radial velocity around voids, given by 5.12, is a simple model of the velocity using a pure geometrical interpretation. In this thesis, I will implement and test whether adding a correction term to equation 5.12, as proposed by [4], will provide a better fit to the data. The proposed model adds a correctional term to the velocity given as

$$v(r) = -\frac{1}{3} H f a r \Delta_v(r) + \Gamma \frac{6}{7} f a H \frac{r}{6} \Delta_v^{(2)}, \quad (5.13)$$

where $\Gamma = +1$, when $\Delta_v(r) > 0$ and $\Gamma = -1$, when $\Delta_v(r) < 0$. In addition to the mass density contrast for voids $\Delta_v(r)$ defined in equation 5.11, an additional $\Delta_v^{(2)}$ is introduced as

$$\Delta_v^{(2)}(r) \equiv \frac{3}{r^3} \int_0^r x^2 \delta(x)^2 dx. \quad (5.14)$$

This semi-empirical term is motivated by including a modified second order Eulerian perturbation term in velocity (eq. (43) in [55])

$$\nabla \cdot v^{(2)} = \frac{3}{7} f_2 a H (\delta^{(1)})^2 + \text{Other terms.} \quad (5.15)$$

This modification to the velocity model will be tested by fitting parameters with data when calculating the cross-correlation function for voids and galaxies $\xi_{vg}^s(\mu, s)$ derived later in section 5.3.3.

5.2 Numerical measurement of correlation functions and multipole expansions.

As mentioned in section 2.1.6 correlation functions are useful tools for measuring the statistical properties of the universe. In order to test analytical models, one has to compare with data from observations or simulations. The code ¹ used for this project to measure angular correlation functions is called pyCUTE. This code is a Python wrapper for a modified version of the CUTE[56] code. The code implements a Landy-Szalay estimator [57] for estimating the angular correlation function. The Landy-Szalay estimator for the angular cross-correlation function, where s denotes redshift space, reads

$$\xi^s(s, \mu) = \frac{\langle D_1 D_2 \rangle - \langle D_1 R_2 \rangle - \langle D_2 R_1 \rangle + \langle R_1 R_2 \rangle}{\langle R_1 R_2 \rangle}. \quad (5.16)$$

In this expression, the terms in the angle brackets denote the number of pairs of the given quantities. In this analysis, subscript 1 denotes void centers and subscript 2 denotes halo positions. D_1 and D_2 refers to the actual data while R is samples from a random distribution resembling the dataset. Each quantity in the squared brackets denote the mean value of a pair with separation r and μ , where $\mu = \cos(\theta)$ is the cosine of the angle between the separation vector and the line of sight direction. However, in this project, I will only deal with simulation data in which includes a complete catalogue without empty patches. The numerical calculation of the correlation function will then consist of simply counting void-halo pairs with separation s and angle μ . An illustration of a pair in the coordinate system of the observer is illustrated in figure 5.3

In real space, due to isotropy and homogeneity, there is no angular dependence on the correlation function and $\xi^r(r)$ is only a function of radius. However, the redshift space distortions themselves are responsible for the angular dependency in $\xi^s(\mu, s)$. Since redshift space distortions are only present when considering the component of a

¹The code for measuring correlation functions utilized for this analysis can be found at <https://github.com/seshnadathur/pyCUTE>.

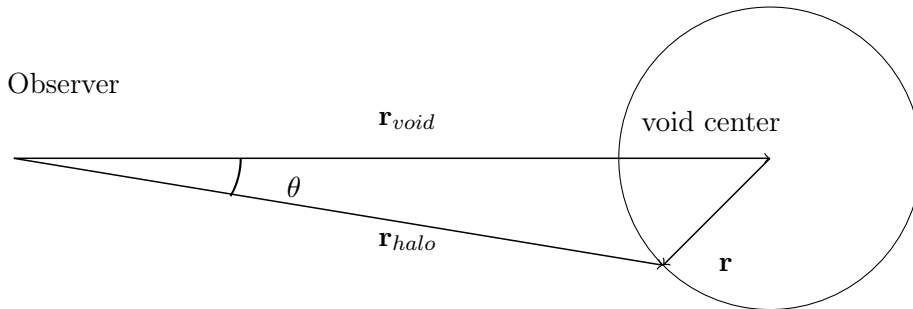


Figure 5.3: Figure illustrating the system used in this analysis. This figure is drawn in real space thereby all vectors denoted as \mathbf{r} . Transformation to redshift space will distort the halo positions and the corresponding system will have vectors denoted as \mathbf{s} .

vector parallel to the line of sight direction, isotropy is not conserved in redshift space. In order to characterize the angular dependencies, the angular cross-correlation function is expanded into multipoles[17]

$$\xi_\ell^s(s) = \int_0^1 \xi^s(s, \mu)(1 + 2\ell)L_\ell(\mu)d\mu, \quad (5.17)$$

where $L_\ell(\mu)$ are Legendre polynomials of order ℓ . The redshift space distortions themselves are fully characterized by the monopole and quadrupole terms[58]. The Legendre polynomials for $\ell = 0$ and $\ell = 2$ are given by

$$L_0(\mu) = 1 \quad (5.18)$$

and

$$L_2(\mu) = (3\mu^2 - 1)/2. \quad (5.19)$$

5.3 Void Analysis

5.3.1 Density profile

After running the void finding algorithm from REVOLVER on our dataset, we have a list of x , y and z coordinates for every void center. With a catalogue of void centers, one has to compare their position to the positions of halos found in their respective catalogue. This allows one to calculate the average density profile as a function of distance from the void center. Using a periodic kd-tree module² to search for the halos nearest to a particular void center with a gradually increasing radius, one can count the halos in the catalogue close to a given void center. For every void, a search for neighbouring halo particles is conducted with a gradually increasing radius. For a linearly spaced radius array $\mathbf{r} = \{r_0, \dots, r_n\}$, using an iterative approach, every halo in a sphere, centered

²The periodic kd-tree code applied for this analysis is found at at https://github.com/patvarilly/periodic_kdtree

around a given void, with radius r_i is counted. The total number of halos is stored for every iteration. For the first iteration, the number of halos in a sphere around the void center is calculated and stored. The density for this sphere is then n_i/V_i , where n_i is the number of halos inside the sphere and $V_i = 4\pi r_i^3/3$. For the next iteration, the number of halos inside a sphere with radius r_{i+1} is counted. A lot of the halos inside this new sphere overlaps with the previous sphere. Therefore the number of halos in the previous sphere was stored and then subtracted from the number of galaxies in the new sphere. This leaves only the number of halos counted in the spherical shell that does not overlap between the two spheres. The volume of this spherical shell is $V_{i+1} = 4\pi(r_{i+1}^3 - r_i^3)/3$. The number density in the shell, defined by subtracting the previous sphere from the new sphere, is the number of galaxies in the new sphere subtracted by the number of galaxies in the previous sphere divided by the volume of the spherical shell. The total number of galaxies in the new sphere is stored and this process is repeated for all iterations defined by the radius vector \mathbf{r} . This process will give the density profile as a function of radius for a given void. By adding the density profiles for all voids and dividing by the number of voids one will get an average density profile for the whole dataset. From the relation for the overdensity, given by equation 2.33, the stacked overdensity profile for the voids in a catalogue is then given by the relation

$$\delta(r) = \rho(r)/\bar{\rho} - 1, \quad (5.20)$$

where $\bar{\rho}$ is the average density in the simulation volume N/V , where N is the total number of halo particles in the halo catalogue and V is the total volume, which in this case is $2500^3 [\text{Mpc}^3/h^3]$. The number of particles for the datasets tested is given in table 4.2.

5.3.2 Velocity profile and velocity dispersion

When studying linear theory applied to voids and filaments, one needs an observational or numerical representation to test its validity. As explained in section 5.1 the radial velocity profile can be represented analytically and may provide a good comparison. In addition to the x , y and z coordinates for the halo particles, the galaxy catalogue used also contains the v_x , v_y and v_z particles. The radial velocity profile is calculated in a similar manner as the density profile for voids described in 5.3.1. Using a periodic kd-tree centered at the current void one can count all particles in a sphere around the current void center. For a given halo particle its position with respect to a given void center is given by

$$\mathbf{r} = \mathbf{r}_{\text{void}} - \mathbf{r}_{\text{halo}}. \quad (5.21)$$

This is the vector that points radially from the void center to a given halo particle. The radial velocity component is the component of the velocity \mathbf{v} for a given halo particle that points along this vector. The component of the vector \mathbf{v} along the radial vector \mathbf{r} is calculated as

$$v_r = \frac{\mathbf{r} \cdot \mathbf{v}}{|\mathbf{r}|}. \quad (5.22)$$

This is the radial velocity for a halo particle with respect to a given void center. In a similar manner to the density profile, the radial velocity is calculated iteratively for every particle in a sphere around the void center. In order to get the radial velocity profile the area around the voids are divided into shells. For every iteration, for a given void, the radial velocity for each sphere is stored for the next iteration. The radial velocity in a given shell is the radial velocity in the new sphere subtracted by the radial velocity in the previous sphere. For each shell, this is averaged by dividing by the number of particles in a given shell. After this has been done for all voids, the radial velocity profile is divided by the number of voids in the catalogue to get the average radial velocity profile for voids in the catalogue.

The velocity dispersion σ_v is also calculated. The velocity dispersion is given by the usual expression for the standard deviation

$$\sigma_v = \sqrt{\langle v^2 \rangle - \langle v \rangle^2}. \quad (5.23)$$

This velocity is along the line of sight for the observer. Since I am working with data from simulations, the observer is located in the xy -plane and all velocities are measured in the z -direction. In real observations the observer would be located at a point and an angular dependency would be introduced. Therefore in this analysis, only the v_z component is considered. This is calculated in spherical shells around each void center to get the average void velocity dispersion profile.

5.3.3 Void-galaxy cross correlation function.

As mentioned in section 2.1.6 correlation functions are essential tools in cosmology for measuring the statistical properties of observables in the universe. In this analysis, the cross-correlation function that will be looked into is the cross-correlation between void centers and dark matter halos. Since dark matter halos in practice traces the galaxy distribution, as galaxies are often located in the center of dark matter halos, this particular correlation function will be referred to by its usual name, namely the void-galaxy cross-correlation function. Following the approach of [17], the following subsection will describe how to derive the void-galaxy cross-correlation function.

Let \mathbf{r}_{void} denote the position of a void center and \mathbf{r}_{halo} denote the position of a halo or galaxy particle in real space. Their separation vector \mathbf{r} is then given by equation 5.21. The system is illustrated in figure 5.3. The relation from redshift space to real space is given as

$$\mathbf{s} = \mathbf{r} + \frac{\mathbf{v}_{pec} \cdot \hat{\mathbf{r}}_{void}}{aH} \hat{\mathbf{r}}_{void}, \quad (5.24)$$

where \mathbf{v}_{pec} is the peculiar velocity of a galaxy or halo particle and $\hat{\mathbf{r}}_{void} = \mathbf{r}_{void}/|\mathbf{r}_{void}|$. Here it is assumed that the velocity of void centers are negligible. This is partially due the modelling of the velocity of the void centers being complicated. It is also assumed that the positions of the void centers themselves are invariant under transformation from real space to redshift space. Since, for this project, data from simulations are utilized one has access to both a catalogue with real space and redshift space positions. In redshift space the distribution of galaxies or halos gets distorted and therefore the

number of voids in real space may not be preserved in redshift space. If one utilizes observational data, only the redshift space positions are observed. Therefore reconstruction, described here in section 3.2.2, has to be applied to regain the field in real space from redshift space. By using the real space and redshift space catalogues provided by the MultiDark simulations[22], as I will do in this thesis, one eliminates errors introduced by the additional step of performing reconstruction on the redshift space data. Cross correlating real space voids with the redshift space galaxy catalogue will satisfy the conditions from the abovementioned assumptions. However, for future work, it is important to study the effect of performing this extra step on simulation data to better quantify the potential error introduced as this step has to be done for this method to be applied on observational data.

The basic assumption when calculating the cross-correlation function required is that the number of void-galaxy pairs in the simulation volume is unchanged during the transformation from real space to redshift space. This is stated mathematically as

$$(1 + \xi_{\text{vg}}^s(\mathbf{s}))d^3s = (1 + \xi_{\text{vg}}^r(\mathbf{r}))d^3r, \quad (5.25)$$

where superscript r and s denotes real space and redshift space respectively. Here $\xi_{\text{vg}}^s(\mathbf{s})$ is the cross correlation between real space voids and the redshifted galaxy field and $\xi_{\text{vg}}^r(\mathbf{r})$ is the cross correlation between real space voids and real space galaxy positions. \mathbf{r} and \mathbf{s} are vectors pointing from the void center to a given galaxy position in real space and redshift space respectively. Another assumption made is that the velocity of all halo particles near a void can be approximated by its radial velocity $\mathbf{v} = v_r \hat{r}$. We orient our coordinate system so that $\hat{r}_{\text{void}} = \hat{z}$. The observer will then only experience redshift distortions in the z direction. This makes $s_x = r_x$ and $s_y = r_y$. With $\hat{r}_{\text{void}} = \hat{z}$, we then have that $\hat{r}_{\text{void}} \cdot \hat{r} = z/r$. We can then rewrite equation 5.24 as

$$\mathbf{s} = \mathbf{r} + \frac{v_r z}{aHr} \hat{z}. \quad (5.26)$$

Here z is the z -component of \hat{r} . The coordinate transformation will then have the following jacobian determinant

$$\left| J\left(\frac{\mathbf{s}}{r}\right) \right| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{\partial s_z}{\partial r_z} \end{vmatrix}, \quad (5.27)$$

where

$$\frac{\partial s_z}{\partial r_z} = 1 + \frac{v_r}{raH} + \frac{(v_r' - v_r/r)}{aH} \mu^2. \quad (5.28)$$

Here v_r' denotes derivative with respect to r . One can now rewrite equation 5.25 as

$$1 + \xi_{\text{vg}}^s(\mathbf{s}) = (1 + \xi_{\text{vg}}^r(\mathbf{r})) \left[1 + \frac{v_r}{raH} + \frac{(v_r' - v_r/r)}{aH} \mu^2 \right]^{-1}. \quad (5.29)$$

This equation can be calculated using both the velocity derived from linear theory for voids, given by equation 5.12, and the modified velocity proposed by [4], given in equation 5.13.

Using the radial velocity v_r for halos derived from linear theory, given in equation 5.12, it is sometimes customary to approximate the expression for the cross correlation function given in equation 5.29. In addition to the radial velocity v_r , we have $\Delta'_v(r) = 3(\delta(r) - \Delta_v(r))/r$ giving

$$v'_r = -Hfa(\delta(r) - \Delta_v(r)), \quad (5.30)$$

where $\Delta_v(r)$ is the density contrast defined in equation 5.11. Inserting this into equation 5.29 yields

$$1 + \xi_{\text{vg}}^s(\mathbf{s}) = (1 + \xi_{\text{vg}}^r(\mathbf{r})) \left[1 - \frac{f}{3}\Delta(r) - f\mu^2(\delta(r) - \Delta(r)) \right]^{-1}. \quad (5.31)$$

This equation can be expanded using the Taylor expansion $1/(1-x) \approx 1+x + \mathcal{O}(x)^2$. When expanding this equation it is appropriate with a review of the terms. It is assumed that for voids the density profile is small enough so that terms of second order in δ and Δ_v is negligible meaning δ^2 , Δ_v^2 and $\delta\Delta_v$ terms are not considered. This leaves us with terms of order δ and $\delta\xi^r$. Equation 5.31 then reduces to

$$\xi_{\text{vg}}^s(s, \mu) = \xi^r(r) + \frac{f}{3}\Delta_v(r)(1 + \xi^r(r)) + f\mu^2[\delta(r) - \Delta_v(r)](1 + \xi^r(r)). \quad (5.32)$$

This is an approximation to equation 5.29. Although the approximations give the model sufficient validity on a range of scales [18], testing during the development of the methodology showed that the full expression given by equation 5.29 proved to correspond better with numerical calculations described in section 2.1.6. Since the equations are implemented numerically there should therefore be no reason to not use the full expression for the sake of this analysis.

As mentioned earlier, it is worth noting that since $\xi^r(r)$ is the correlation function in real space coordinates, if one is to apply this method to observational data, reconstruction, described in section 3.2.2, has to be performed. Since observational data is observed in redshift space, one cannot observe $\xi^r(r)$ directly. One therefore has to perform reconstruction on the data and then perform the correlation function calculation, described in section 5.2, in order to obtain $\xi^r(r)$. When working with simulation data one possesses both the real space and redshift space catalogues. In this analysis $\xi^r(r)$ is therefore calculated by cross-correlating the real space catalogue with real space voids using pyCUTE and no reconstruction is performed.

The correlation function in redshift space $\xi_{\text{vg}}^s(\mathbf{s}, \mu)$ relies on the real space length of the separation vector $|\mathbf{r}|$, as can be seen from equations 5.32 and 5.29. For every distance from the void center s in redshift space, we want the corresponding length from the void center r in real space to calculate $\xi_{\text{vg}}^r(r)$, $\delta(r)$ and $\Delta_v(r)$. To achieve this we start by decomposing s into a parallel and perpendicular component using basic trigonometry as

$$s_{\parallel} = s\mu \quad (5.33)$$

and

$$s_{\perp} = s(1 - \mu^2). \quad (5.34)$$

In the simulation volume, there are only redshift space distortions along the line of sight direction. This will mean that the perpendicular component is unchanged when changing coordinates from redshift space to real space. The parallel component is found by inserting equation 5.12 into equation 5.24 and using the fact that there are no distortions in the perpendicular direction as

$$r_{\parallel} = s_{\parallel} + s \frac{\beta \Delta(s) \mu}{3}, \quad (5.35)$$

where $\beta = f/b$. The corresponding real space separation is then given as

$$r = \sqrt{r_{\parallel}^2 + r_{\perp}^2}. \quad (5.36)$$

5.3.4 Expanding the model to account for velocity dispersion

The previous model for the void-galaxy cross-correlation function is a good approximation. However, it can be improved by taking into account the velocity dispersion of galaxies around the void centers in the line of sight direction. This will provide a more accurate model thereby giving a better fit to data. Instead of only considering radial velocities, in addition, another component along the line of sight direction is added

$$\mathbf{v} = v_r \hat{r} + v_{\parallel} \hat{r}_{void}. \quad (5.37)$$

v_r is the radial velocity used in the previous model given by equation 5.12. v_{\parallel} is given by a gaussian probability distribution $P(v_{\parallel})$ with zero mean

$$P(v_{\parallel}) = \frac{1}{\sqrt{2\pi}\sigma_{v_{\parallel}}} \exp\left[-\frac{v_{\parallel}^2}{2\sigma_{v_{\parallel}}^2}\right]. \quad (5.38)$$

Since we in our simulation volume only measure velocities along the line of sight direction, we can take $\sigma_{v_{\parallel}}$ to be measured by equation 5.23. If this method is to be used on observations, one cannot necessarily extract the velocities of each particle and gain information about σ_v . Therefore simulations have to be utilized in order to calculate this quantity. This will however introduce free parameters when doing parameter estimations as this profile has to be scaled differently when derived from another dataset. Using equation 5.38, one will get an integral for calculating the improved model taking into account dispersion in the velocity of halos [18]

$$1 + \xi^{s,streaming}(s, \mu) = \int_{-\infty}^{\infty} \frac{1 + \xi^{s,base}(\mathbf{s})}{\sqrt{2\pi}\sigma_{v_{\parallel}}(r)} \exp\left[-\frac{v_{\parallel}^2}{2\sigma_{v_{\parallel}}^2(r)}\right] dv_{\parallel}. \quad (5.39)$$

In this equation, $\xi^{s,base}(s)$ can be represented by equations 5.32 or 5.29, dependent on whether one uses the aforementioned approximations or not. In this thesis, I will use the full expression given by equation 5.29.

5.4 Filament Analysis

5.4.1 Masking the simulation volume

When applying the DisPerSE code on the particle distribution the output will be, for each filament detected, coordinates to the start point of each segment making up the total filament. For each filament there will be a number of filament segments with start points \mathbf{r}_i . This also makes \mathbf{r}_{i+1} the endpoint of filament segments with start points \mathbf{r}_i . Each filament segment F_j is then given by

$$\mathbf{F}_j = \mathbf{r}_{i+1} - \mathbf{r}_i. \quad (5.40)$$

Here j denotes a given filament segment while i denotes the start points of the segments. With this notation $j + 1$ is denoted by start points $i + 2$ and $i + 1$ and so on. When a given filament segment is defined the simulation volume is masked to remove all particles outside the edges of the start and endpoint of the filament segment. This is done by defining two planes perpendicular to \mathbf{F}_j with points \mathbf{r}_{i+1} and \mathbf{r}_i . The equation for a plane P given a vector \mathbf{F} and a point \mathbf{r} is given by the usual formula

$$P = F_x(x - r_x) + F_y(y - r_y) + F_z(z - r_z) = 0, \quad (5.41)$$

where x , y and z are coordinates of the particles in the simulation volume. In this section, this vector will be denoted as \mathbf{x} . The coordinates r_x , r_y and r_z are the components of \mathbf{r} . For a given filament segment F , with start point r_i and end point r_{i+1} , the particles kept are the particles satisfying the following criteria

$$P_0 = F_x(x - r_{i_x}) + F_y(y - r_{i_y}) + F_z(z - r_{i_z}) > 0 \quad (5.42)$$

and

$$P_1 = F_x(x - r_{i+1_x}) + F_y(y - r_{i+1_y}) + F_z(z - r_{i+1_z}) < 0. \quad (5.43)$$

The rest of the particles are masked keeping only the particles in the simulation volume bound by the two plane parallel to the filament segment centered in the start point and endpoint. A figure of this process is illustrated in figure 5.4

5.4.2 Density profile

After masking the particle box, as described in section 5.4.1, we are left with a filament segment \mathbf{F}_j defined by the start and endpoints \mathbf{r}_i and \mathbf{r}_{i+1} . The distance from each halo particle in the masked simulation volume to the filament segment is the length of the vector from the particle perpendicular to the filament axis

$$d = \frac{|\mathbf{x} \times \mathbf{F}_j|}{|\mathbf{F}_j|}. \quad (5.44)$$

This will leave us with a catalogue of the closest distance for all particles to the axis of the filament segment. A histogram is then created of this catalogue to easily count

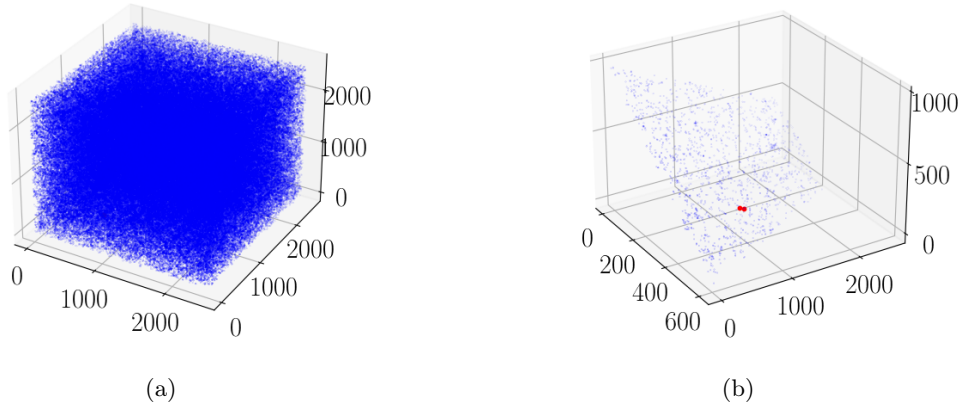


Figure 5.4: Figure illustrating visually how the halo catalogue is masked. On the left is a scatter plot of the halo catalogue without a mask. The right figure illustrates the masking of the catalogue where the two red dots are the start- and end-points of the filament segment. Here one can see that only particles bound by the spine of the filament segment itself are kept. The dataset used for this illustration is the MD1 dataset with $\sigma = 1$ given as persistence threshold to DisPerSE. All units on the axes are in Mpc/h.

the number of particles in a given radial distance bin from the filament segment. The volume of the spherical shells surrounding the filament segment containing the particles in the radial bin is then given as

$$V_i = \pi |\mathbf{F}_j| (r_i^2 - r_{i-1}^2), \quad (5.45)$$

where r_i and r_{i-1} denotes the edges of the radial bins for the histogram. The radial density profile of the filament segment is then given as

$$\rho_i = \frac{N_i}{V_i}, \quad (5.46)$$

where N_i is the number of particles in each radial bin. For each filament segment, the density is added together and divided by the number of segments to get the density profile of a given filament. After adding the average density for all filaments, it is then divided by the total number of filaments in the catalogue in order to get the average density profile.

5.4.3 Velocity profile

In the catalogue of halo particles, the v_x , v_y and v_z components of the halo particle velocity \mathbf{v} is given. In order to validate the efficacy of the velocity model derived from linear theory, given in equation 5.7, one needs data to compare with the model. This is done by calculating the velocity component perpendicular to the filamentary spine.

By averaging over all filaments, one gets the stacked radial velocity profile. As for the calculation of the density profile in section 5.4.2, let $\mathbf{F}_j = \mathbf{r}_{i+1} - \mathbf{r}_i$ define a given filament segment. After, for each filament, the particle box is masked as described in section 5.4.1, the radial component d perpendicular to the filament axis is calculated. This component is given by equation 5.44. The length of the parallel component of the halo positions \mathbf{x}_{\parallel} is given by the Pythagorean theorem

$$|\mathbf{x}_{\parallel}| = \sqrt{|\mathbf{x}|^2 - d^2}. \quad (5.47)$$

With the length of the parallel component calculated, one gets the parallel component of the vector \mathbf{x} itself by multiplying with the normal vector for the filamentary spine

$$\mathbf{x}_{\parallel} = |\mathbf{x}_{\parallel}| \frac{\mathbf{F}_j}{|\mathbf{F}_j|}. \quad (5.48)$$

This allows one to easily calculate the component perpendicular to the filamentary spine

$$\mathbf{x}_{\perp} = \mathbf{x} - \mathbf{x}_{\parallel}. \quad (5.49)$$

To obtain the component of the velocity perpendicular to the filamentary spine, one can use the regular formula for calculating the component of a vector along another vector

$$v_r = \frac{\mathbf{v} \cdot \mathbf{x}_{\perp}}{|\mathbf{x}_{\perp}|}. \quad (5.50)$$

This is calculated for all particles in a given radial bin, as defined by the histogram created in the calculation of the density profile, described in section 5.4.2. For all filamentary segments, the velocity in each bin is added together. The same is done for the number of halo particles in each bin. To get the average stacked velocity profile for filaments, the total velocity is divided by the number of particles in each respective bin. This approach in practice gives the stacked velocity profile for filament segments.

5.5 Parametrizing distance scales.

In order to implement our model, we have to introduce a fiducial cosmology. We have to assign values for parameters such as Ω_m and Ω_{Λ} for the model to function. These fiducial values we assign however may differ from the true cosmology. This will introduce distortions in the distance measurements. The Alcock-Paczynski effect[59] provides a test that allows for calibration of the model. In the correct cosmological model, one expects cosmological phenomena, such as clustering of galaxies, to appear isotropic. Observed anisotropies may be attributed to the conversion of distance parameters using the wrong fiducial cosmology. Observed distances perpendicular to the line of sight direction scale proportional to the angular diameter distance $d_A(z)$. Distances parallel to the line of sight direction scale proportional to $1/H(z)$. This allows for the calibration of anisotropic distortions introduced by the fiducial cosmological model itself. As

conducted by [18], the method applied to the analysis in this project will also parametrize these distortions by introducing two dimensionless ratios. One in the perpendicular and one in the parallel direction as

$$\alpha_{\perp} = \frac{d_A(z)}{d_A^{\text{fid}}(z)} \quad (5.51)$$

and

$$\alpha_{\parallel} = \frac{H^{\text{fid}}(z)}{H(z)}, \quad (5.52)$$

where d_A is the angular diameter distance. The correlation function $\xi^s(s, \mu)$ is measured using the fiducial components $s_{\parallel}^{\text{fid}}$ and s_{\perp}^{fid} . These components are then rescaled using the dimensionless ratios as

$$s_{\parallel} = \alpha_{\parallel} s_{\parallel}^{\text{fid}} \quad (5.53)$$

and

$$s_{\perp} = \alpha_{\perp} s_{\perp}^{\text{fid}}. \quad (5.54)$$

Likewise the correlation function in real space $\xi^r(r)$ takes into account distances in real space at a given redshift for a fixed cosmology. The corresponding distance in the real universe may vary from this distance and therefore the real space distance is rescaled as

$$r = \int_0^1 \alpha_{\parallel} r^{\text{fid}} \sqrt{1 + (1 + \mu^2) \left[\frac{\alpha_{\perp}^2}{\alpha_{\parallel}^2} - 1 \right]} d\mu. \quad (5.55)$$

This rescaling of parameters gives an additional parameter when fitting the models to the data as both α_{\perp} and α_{\parallel} has to be tested in order to determine the best fit for the parametrization of the distance scaling. In simulations, the fiducial cosmology is known and $\alpha_{\parallel} = \alpha_{\perp} = 1$. Therefore constraining $\epsilon = \alpha_{\perp}/\alpha_{\parallel}$ for a model applied to simulation data can provide a good test on how the model performs. These parameters are one of the most important parameters of the model as these are directly linked to the cosmological parameters Ω_{Λ} and Ω_m . In [3], the authors show that the study of voids is important additions to other established cosmological surveys for constraining the cosmological parameters. This is shown in figure 5.5

5.6 Calibrating dark matter profiles.

When using a linear bias approximation, we can simply approximate $\delta_g(r) = \xi_{\text{vg}}^r(r)$, as the real space correlation function between voids and galaxies and the normalized density profile $\delta_g(r)$ effectively traces the same quantity. The linear bias approximation itself is to just use $\delta_{\text{dm}} = \delta_g/b$. However, when using a dark matter density profile the bias factor is not considered and instead σ_8 is used to scale the amplitude of the density profile. At a given redshift z we then have

$$\delta_{\text{dm}}(r, z) = \frac{\sigma_8}{\sigma_8^{\text{MD}}} \delta_{\text{dm}}^{\text{fid}}(r, z), \quad (5.56)$$

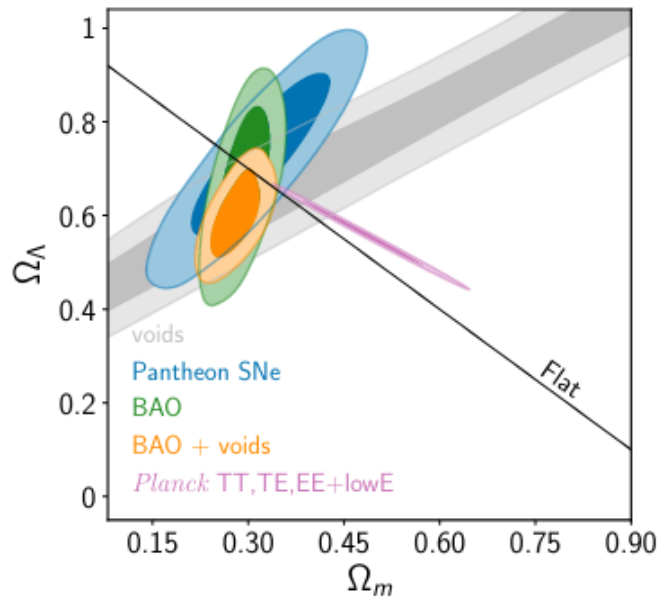


Figure 5.5: Marginalized constraints on Ω_m and Ω_Λ inferred from different cosmological surveys. This figure illustrates that the study of voids is a good addition to other experiments in contributing to providing constraints for these cosmological parameters. Figure is copied from [3].

where $\sigma_8^{\text{MD}} = 0.8228$ is the value for σ_8 for the MultiDark simulation dataset with $z = 0$. For $\delta_{dm}^{\text{fid}}(r, z)$, the exact dark matter density profile for the four simulated datasets was not obtained. Instead, I will use a dark matter profile for voids obtained from the BigMD mock CMASS galaxy sample [60]. This dark matter profile will need some extra scaling in order for it to accurately represent the dark matter profile for the dataset used in this analysis. Its amplitude is scaled by σ_8 , as shown in equation 5.56. However, before the analysis, I will scale the dark matter profile using eye measurements. This is in order to make sure its amplitude is approximately equal to $\delta_g(r)$ derived from the dataset. After this, I will divide by the bias factor for the given dataset in order to gain an approximate amplitude for $\delta_{dm}(r)$. This will make the amplitude roughly in line with what the corresponding dark matter profile for the dataset would have looked like. However, the slope of $\delta_{dm}(r)$ may differ. Therefore, when scaling the radius, an extra r_{scale} parameter is multiplied with equation 5.55 when scaling the radius for the dark matter density profile. The r_{scale} parameter will be included in the likelihood analysis as a part of the model when fitting the data. This will make it so that both the amplitude and the slope of the dark matter density profile will be adjusted when fitting the parameters to the data.

5.7 Likelihood analysis.

In order to test the validity of the models and to see whether the proposed improvements perform better, a likelihood analysis is conducted. The likelihood analysis in this thesis largely follows the approach conducted by [18]. The code implementing the Metropolis-hastings MCMC chain is provided by the publicly available Victor code³. Using the method described in section 5.2, the angular cross-correlation function $\xi^s(s, \mu)$ and its multipoles is calculated from a simulated halo catalogue in redshift space and voids located from the corresponding real space catalogue using REVOLVER. For the purpose of this analysis the MultiDark dataset, described in section 4, is utilized. The numerical angular cross-correlation function is measured using 80 linearly spaced bins for $0 < \mu < 1$ and 30 linearly spaced bins for $2\text{Mpc}/h < s < 118\text{Mpc}/h$. Using the same number of bins and data points, the corresponding correlation function in real space $\xi^r(r)$ is also calculated from the real space halo catalogue. Using the methods described in sections 5.3.1 and 5.3.2, the density profile and velocity dispersion are calculated for each void in catalogue identified by REVOLVER. With these quantities calculated one has all the components needed to evaluate equation 5.39.

5.7.1 Estimating the covariance matrix

The covariance matrix used is calculated using patchy mocks provided by the MultiDark simulation[61][62]. The covariance matrix is provided together with the code used for the likelihood analysis. A mock catalogue is a galaxy or dark matter halo catalogue

³The code used for the likelihood analysis is provided by <https://github.com/seshnadathur/victor>

that is built from simulations designed to resemble an observed galaxy catalogue. The creation of the covariance matrix used is described in [18]. The covariance matrix is calculated using 1000 patchy mocks. For each patchy mock the angular-cross correlation is measured and expanded into a monopole ξ_0^s and a dipole ξ_2^s . The multipoles are then combined into a data vector $\xi^s \equiv (\xi_0^s, \xi_2^s)$ for each patchy mock catalogue. The covariance matrix is then calculated as

$$C = \frac{1}{N-1} \sum_{k=1}^N (\xi^k - \bar{\xi}^k)(\xi^k - \bar{\xi}^k)^T, \quad (5.57)$$

where $\bar{\xi}^k$ is the mean over all the mocks used and k is the index for a given mock. N is the number of mocks used to calculate the covariance matrix itself. In this case, we have $N = 1000$.

5.7.2 Fitting parameters to data.

The theory model, given by equation 5.39, relies on the parameters σ_v and $\beta = f/b$ through the modelling of v_r , described in section 5.3.2. Through the rescaling of r , described in section 5.5, it also relies on α_\perp and α_\parallel , in which is represented as $\epsilon = \alpha_\perp/\alpha_\parallel$. Without the linear bias approximation, the model does not rely on the bias factor b . Instead, it relies on the parameters $f\sigma_8$ and r_{scale} , as described in section 5.6. For the linear bias approximation, with and without the added velocity term described in section §5.1.3, the model then relies on the three parameters $(\beta, \sigma_v, \epsilon)$. When using a dark matter profile, the model relies on the four parameters $(f\sigma_8, \sigma_v, \epsilon, r_{\text{scale}})$

The Victor code utilized for the likelihood analysis implements the approach described by Sellentin & Heavens [40] and calculates the χ^2 for each step in the MCMC chain as

$$\chi^2 = (\xi^{s,\text{streaming}} - \xi^s) \mathbf{C}^{-1} (\xi^{s,\text{streaming}} - \xi^s). \quad (5.58)$$

Here $\xi^{s,\text{streaming}}$ is given by equation 5.39. ξ^s is the numerically calculated cross-correlation function using the real space void catalogue and redshift space halo catalogue. The log-likelihood is then calculated as

$$\ln \mathcal{L} \propto -\frac{N}{2} \ln \left(1 + \frac{\chi^2}{N-1} \right). \quad (5.59)$$

Chapter 6

Results/Discussion

The following section is split into two sections. One section is devoted to the analysis conducted on filaments and another section to the analysis conducted on voids. Throughout this analysis, a fiducial cosmology is assumed. Every time a quantity relying on cosmological parameter is illustrated, the following values were used: $H_0 = 100h(\text{km/s})/\text{Mpc}$, $\Omega_m = 0.31$ and $\Omega_\Lambda = 0.69$.

6.1 Filament analysis

6.1.1 Effect of choosing persistence ratio on dataset

As discussed in section 3.1.4, DisPerSE has an option to filter noise through setting a persistence threshold. This has a profound effect on the number of filaments identified. Figures 6.1 and 6.2 shows slices of the MD1 and MD2 datasets. In these figures, filaments identified by DisPerSE are overplotted. The sigma thresholds ranges from no cuts, $\sigma = 1$, $\sigma = 2$ and $\sigma = 3$. From these figures, one can see that increasing the σ threshold will reduce the number of filaments detected. Figures 6.3 and 6.4 shows histograms of the filament lengths for all filaments in the MD1 and MD2 datasets with the different σ thresholds. From the histograms, one can clearly see that filaments with lengths in the range $20\text{Mpc}/h$ to $50\text{Mpc}/h$ are the ones that get significantly reduced as the σ threshold increases. By looking at the scatter plots for the two datasets, one can see that with no cuts, a lot of knots appear in the intersection where filaments meet. This leads to many individual structures identified as filaments that in reality should only make up one individual filament. On the other hand, using $\sigma = 3$, leaves out a lot of structures that should be interpreted as filaments. Therefore, the following analysis will be conducted using the filament catalogues output by disperse using $\sigma = 1$ and $\sigma = 2$. Looking at the histograms, a lot of very large filaments with length $l \geq 100$ is also detected. Although the code is utilized on a dark matter halo catalogue and not a galaxy catalogue, which in principle could make us detect larger structures, filaments with length larger than $100 \text{ Mpc}/h$ is neglected. The existence of very large structures, such as the Hercules-Corona Borealis Great Wall, which is approximated to measure around $2000 - 3000\text{Mpc}$

in diameter [63], has been proposed. Although, to not account for structures larger than what breaks with the cosmological principle, the largest filaments are cut from the catalogue. To avoid small filaments making up potential knots in the intersection where many filaments meet, filaments with length $l \leq 20\text{Mpc}/h$ are also cut from the catalogue.

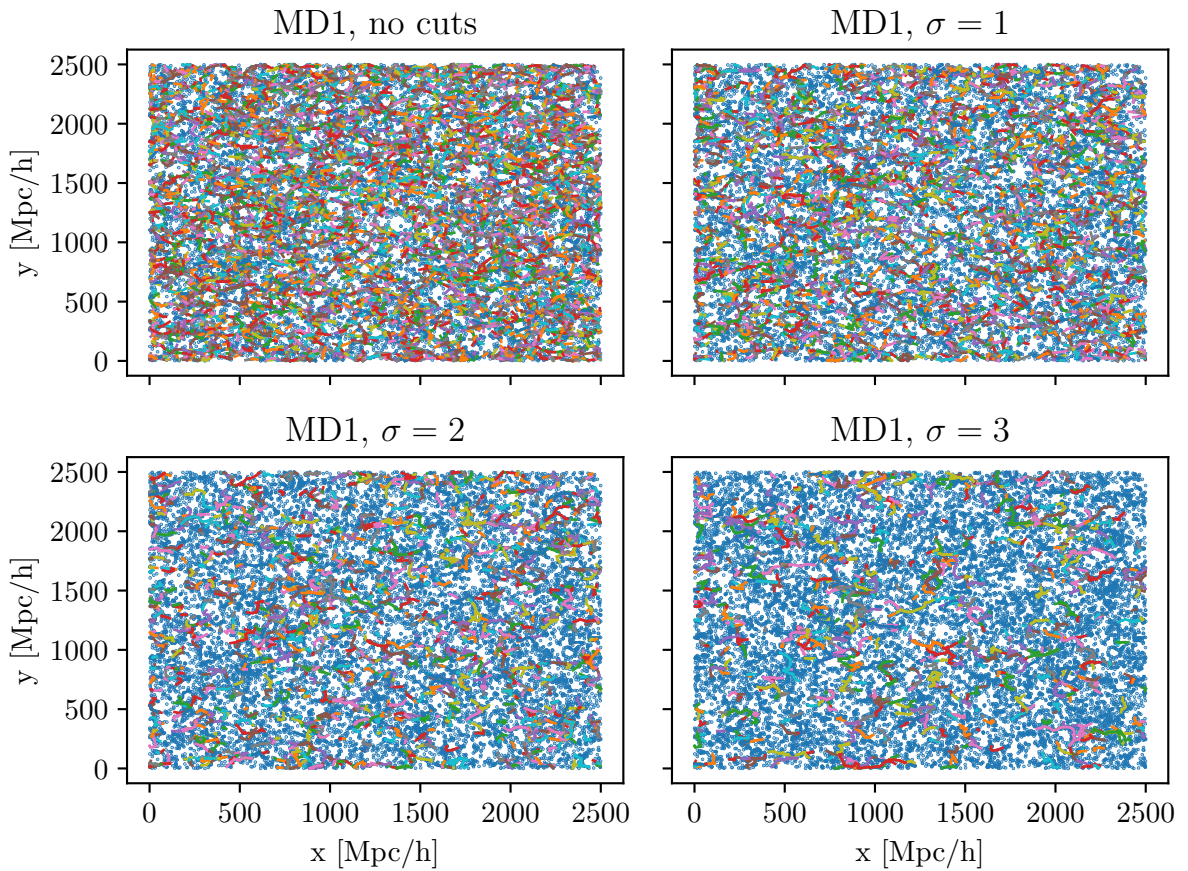


Figure 6.1: Figure showing slices of the MD1 dataset in the xy -plane for $950\text{Mpc}/h \leq z \leq 1050\text{Mpc}/h$. Here the persistence threshold given to DisPerSE is varied between no cuts, $\sigma = 1$, $\sigma = 2$ and $\sigma = 3$. Each blue dot represents a halo particle while all lines represents filaments assigned by DisPerSE.

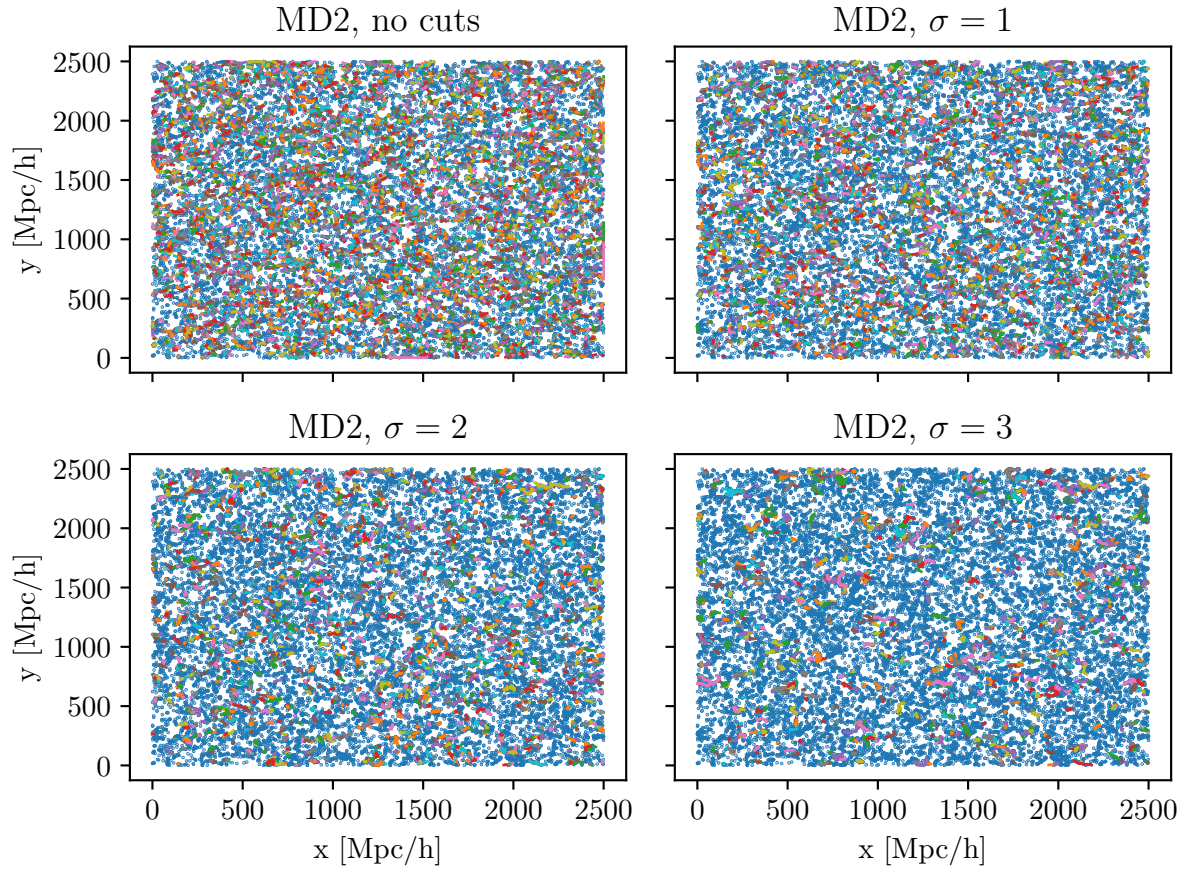


Figure 6.2: Figure showing slices of the MD2 dataset in the xy -plane for $990\text{Mpc}/h \leq z \leq 1010\text{Mpc}/h$. Here the persistence threshold given to DisPerSE is varied between no cuts (a), $\sigma = 1$, $\sigma = 2$ and $\sigma = 3$. Each blue dot represents a halo particle while all lines represents filaments assigned by DisPerSE.

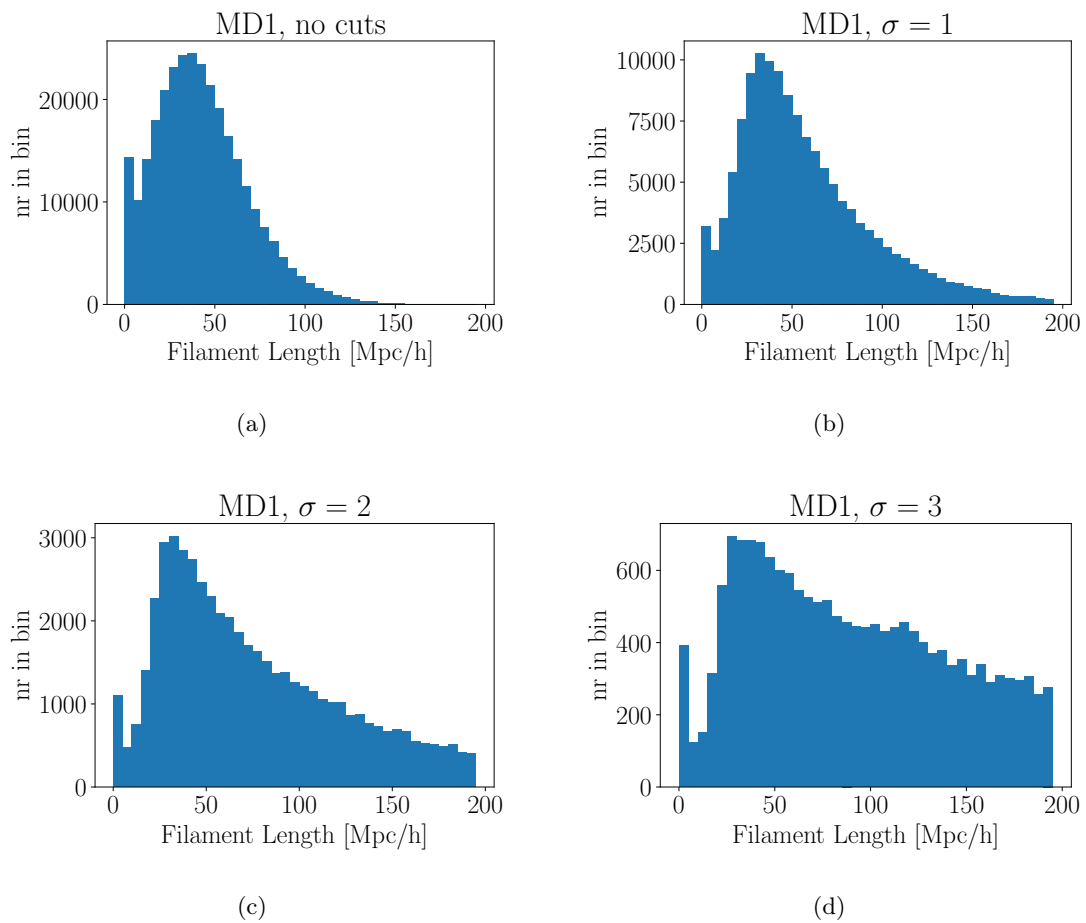


Figure 6.3: Figure showing histograms of filament lengths calculated by disperse for the MD1 dataset. In this figure, the persistence threshold given to DisPerSE is varied between no cuts (a), $\sigma = 1$ (b), $\sigma = 2$ (c), $\sigma = 3$ (d).

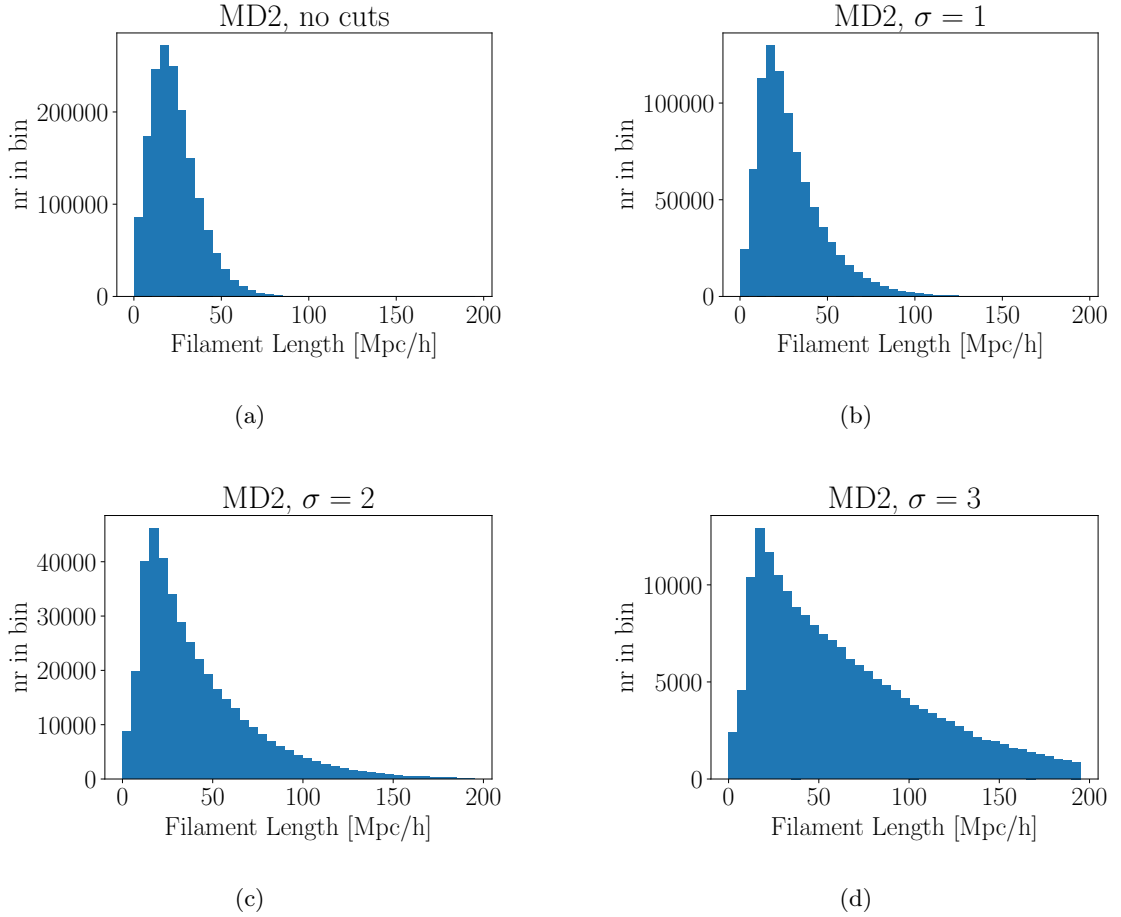


Figure 6.4: Figure showing histograms of filament lengths calculated by disperse the MD2 dataset. In this figure, the persistence threshold given to DisPerSE is varied between no cuts (a), $\sigma = 1$ (b), $\sigma = 2$ (c), $\sigma = 3$ (d).

6.1.2 Density profiles

In accordance with the method described in section 5.4.2, the density profile $\rho_g(r)/\bar{\rho}$ was calculated for the MD1 and MD2 datasets with $\sigma = 1$ and $\sigma = 2$ as inputs to the DisPerSE code. The density profile was calculated from $r = 0\text{Mpc/h}$ to $r = 300\text{Mpc/h}$ using linearly spaced bins of size 2Mpc/h . This is a very long distance, but in order to test the scales at which linear theory can be applied to filaments, an arbitrarily large distance was chosen. From the histogram of filament lengths, illustrated in figures 6.3 and 6.4, one can see that increasing the σ threshold for the DisPerSE code greatly reduces the amount of shorter filaments. From the MD1 dataset for both σ -thresholds chosen, illustrated in figures 6.5 and 6.6, one can see that the density close to the filament spine is larger for the higher σ -threshold. This is also present for the MD2 dataset,

as shown when comparing figures 6.7 and 6.8. As one can see from the histograms of filaments, increasing the σ -threshold reduced the number of smaller filaments. This may suggest that longer filaments possess a higher number density of particles per volume. This result may also be an effect of noise removed when cutting shorter filaments. The shorter filaments, making up a larger portion of the total number of filaments in the lower σ -threshold, may be structures identified in more sparse areas of the simulation volume thereby lowering the stacked density profile.

As can be seen from the figures of the density profiles, a power law model was fitted to the density profile from the filament spine to approximately where the profile reaches the average density of the simulation volume. For the MD1, dataset the density profile flattens out at approximately $r = 30\text{Mpc/h}$ for both σ -thresholds. For the MD2 dataset, this happens at roughly $r = 20\text{Mpc/h}$ for both σ -thresholds. This distance is longer than what is detected by [64] (e.g fig.9). Using a similar method applied to the TNG300 simulation[65] at redshift $z = 0$, they found that the density profile converged to the average density of the simulation volume at approximately $r = 7\text{Mpc/h}$ - 21Mpc/g depending on dataset and length cuts applied to filaments. They also found that long filaments converged faster than shorter filaments. A similar method was also applied by [66] (e.g fig.5) to galaxies in the range $0.1 < z < 0.3$ using the WISExSCOS catalogue[67]. They calculated the stacked overdensity profile and their results show that the gradient flattens out at approximately 21Mpc/h . The calculations in this thesis on the other hand is conducted at $z = 1$. For both datasets and σ -thresholds a power law $\propto r^x$ was fitted using the density profiles from $r = 0\text{Mpc/h}$ to $r = 30\text{Mpc/h}$ for the MD1 dataset and $r = 0\text{Mpc/h}$ to $r = 20\text{Mpc/h}$ for the MD2 dataset. These results suggest that there is a steeper gradient when increasing the σ threshold. This may suggest that longer filaments are denser or it may simply be an artefact of DisPerSE removing shorter filaments that should be perceived as noise. In order to test this, it would be useful for future expansion of these results to test the algorithm on filaments of different lengths by applying manual cuts to the filament catalogue using the same σ -threshold.

After the density profiles reach the average density of the simulation volume, for the two sigma thresholds in the MD1 dataset, there is a prominent bulge that peaks at approximately $r = 100\text{Mpc/h}$. For the density profiles derived from the MD2 dataset, this bulge is not as prominent, but after approximately $r = 60\text{Mpc/h}$, the density profile starts to decline slowly. This effect may be due to a potential artefact where the volume of the cylindrical shells surrounding the filament does not scale properly with the number of halos counted in each shell. However, the prominent bulge in the MD1 dataset may instead suggest that this is a feature of the dataset itself. The MD2 dataset, in which is a lot more dense seems to wash out this bulge as it is not as prominent in this dataset. There is also a prominent difference in that when choosing $\sigma = 1$ or $\sigma = 2$ as input to DisPerSE, the level at which the gradient flattens out relative to the background density changes slightly. This however may be due to increasing σ -threshold effectively lowers the amount of shorter filaments which, again, may suggest that longer filament are more dense. It may also be an effect of DisPerSE correctly rooting out filaments

that should be perceived as noise when increasing σ -threshold. The apparent bulge in the MD1 dataset may make this density profile problematic when applying it to linear theory to calculate the velocity, given in equation 5.7. However, the density profiles for the MD2 dataset should be sufficient on scales up to $r = 60\text{Mpc}/h$.

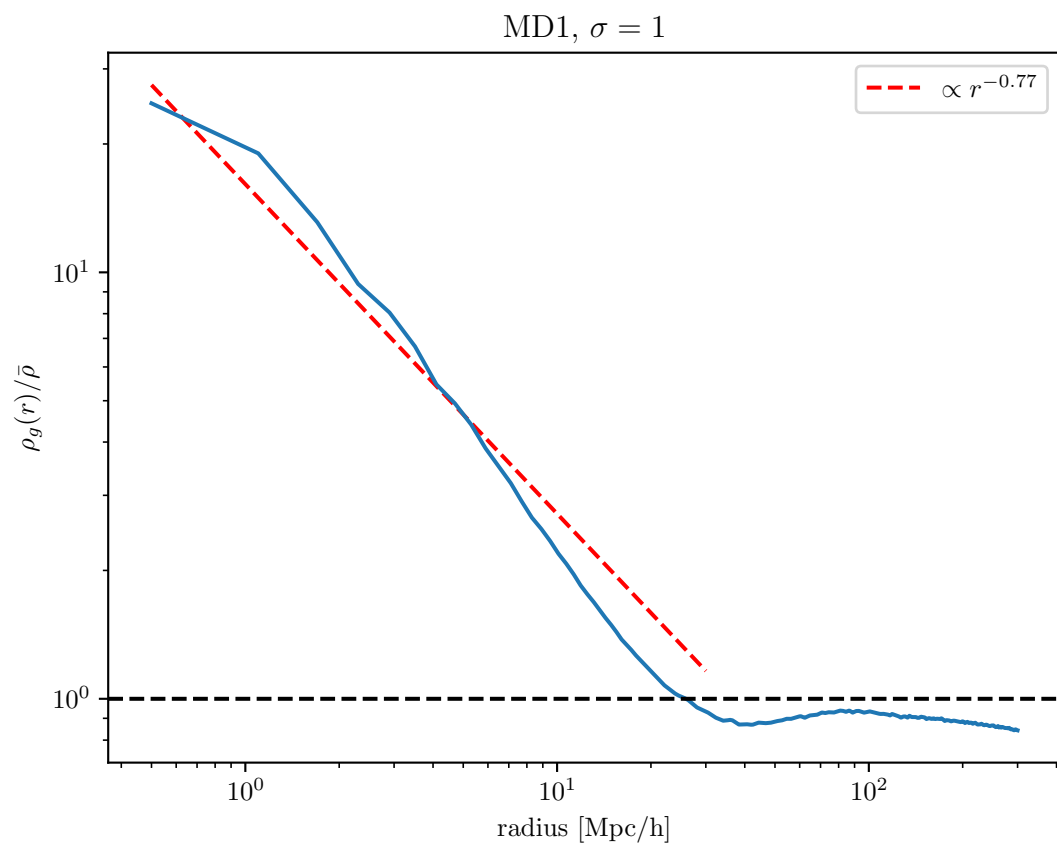


Figure 6.5: Figure showing the calculated density profile $\rho_g(r)/\bar{\rho}$ for filaments in the MD1 dataset with $\sigma = 1$.

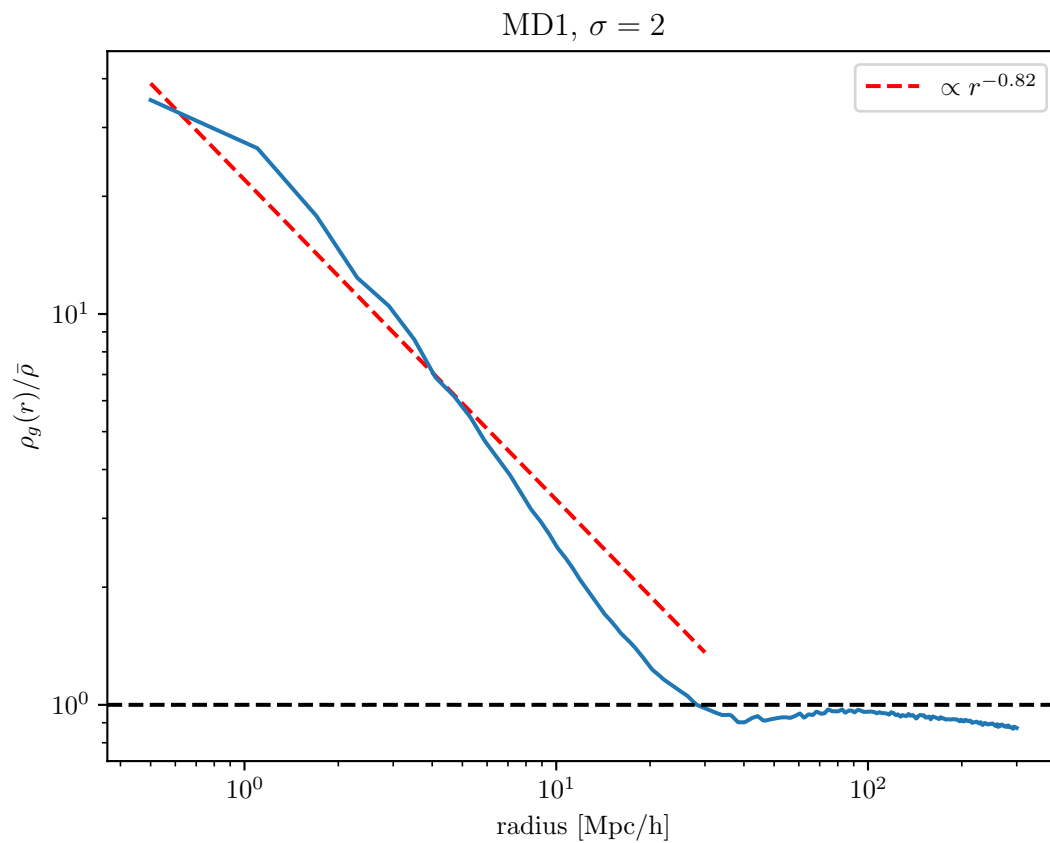


Figure 6.6: Figure showing the calculated density profile $\rho_g(r)/\bar{\rho}$ for filaments in the MD1 dataset with $\sigma = 2$.

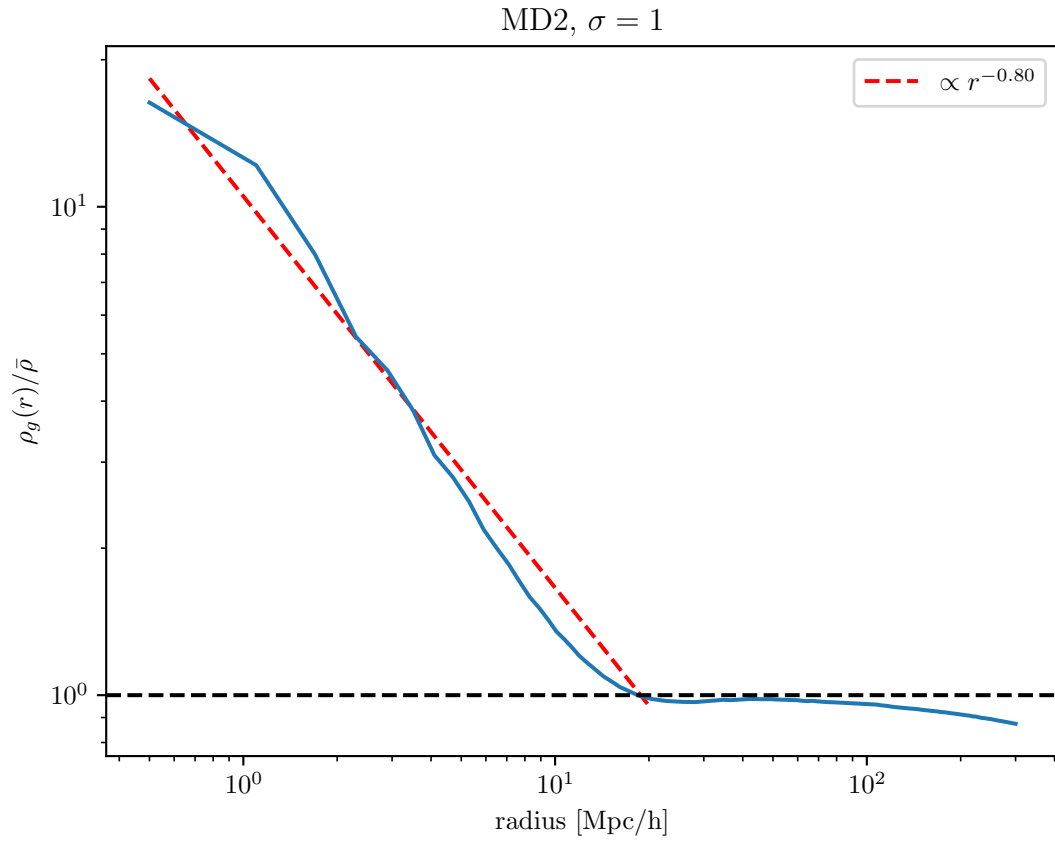


Figure 6.7: Figure showing the calculated density profile $\rho_g(r)/\bar{\rho}$ for filaments in the MD2 dataset with $\sigma = 1$.

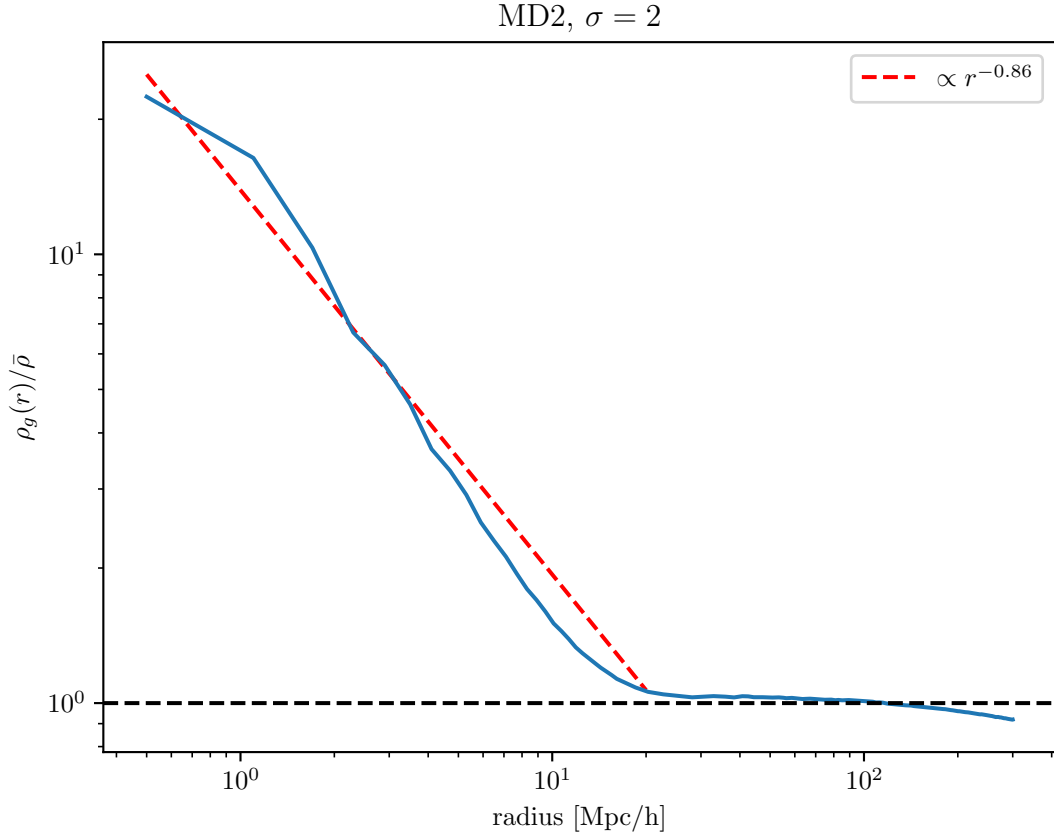


Figure 6.8: Figure showing the calculated density profile $\rho_g(r)/\bar{\rho}$ for filaments in the MD2 dataset with $\sigma = 2$.

6.1.3 Radial velocity profiles

In order to test whether linear theory can be used to approximate the behavior of matter around filaments, the theory, described in section 5.1.1, was applied using the density profiles shown in section 6.1.2. The velocity profile, calculated numerically from the dataset, was calculated from $r = 0\text{Mpc/h}$ to $r = 300\text{Mpc/h}$ using linearly spaced bins of size 2Mpc/h . Figures 6.9 and 6.10 shows the velocity of filaments predicted by linear theory compared with the velocity calculated from the dataset, as described in section 5.4.3, for the MD1 dataset with $\sigma = 1$ and $\sigma = 2$ respectively. Figures 6.11 and 6.12 shows the same for the MD2 dataset. As can be seen from the figures of the density profiles in the previous section, the density profiles did not have a gradient that was approximately zero after it flattened out and reached the average density of the simulation volume. Instead, it showed a slow decline for the MD2 dataset and a more

prominent bulge for the MD1 dataset. The small negative gradient in the density profile causes the velocity predicted by linear theory to have an artificial increase on large scales. In order to counteract this, the overdensity $\delta_g(r) = \rho_g(r)/\bar{\rho} - 1$ was artificially set to zero for scales $r > 50\text{Mpc/h}$. For all datasets, the numerically calculated velocity has a radial component with a relatively large value close to the filament spine before it declines and eventually reaches zero. This is reasonable as far away from the filament spine, the stacked velocity profile should not be biased in any direction. All the velocity profiles display a feature in which appears as a hook close to the filament spine. This may be due to the halos being inside the densest parts of the filament experience a greater gravitational attraction from multiple sides making the net acceleration towards the filament spine smaller. Another explanation for this effect may be that one could imagine the densest parts of the filaments being at the edges of the filaments in large clusters. Overdensities at the end-points of the filament spine will make matter flow along the vector parallel to the spine itself, thereby decreasing the amplitude of the radial component. Studying the two-dimensional velocity field around filaments may be of interest in order to verify whether this could be a reasonable explanation. From the figures, it is also evident that the measured velocity profiles flattens out and starts to converge towards the background density of the simulation volume on scales $r > 100\text{Mpc/h}$.

The velocities predicted by linear theory displays the same behavior as the velocity measured from the catalogues themselves. Their amplitude however is much larger than what should be expected as measured from the dataset. As expected there is also a prominent bend on the velocity profile predicted by linear theory at $r = 50$, where $\delta_g(r)$ was set to zero. Since one has the simulation to compare with, one could use eye measurements to move the radius at which one sets $\delta_g(r) = 0$. However, if this was to be used for observational data, one is not able to make this comparison, and therefore, this limit was set at $r = 50\text{Mpc/h}$. As can be seen from the density profiles from the previous section, choosing $\sigma = 2$, made the measured density at the center of the filament larger. This is also evident in that the velocity profiles predicted by linear theory has a higher amplitude for $\sigma = 2$ relative to $\sigma = 1$ close to the filament spine. Although the density profiles using $\sigma = 2$ gave more reasonable results for large radii, since $\delta(r > 50\text{Mpc/h})$ is set to zero, this potential effect on the predicted velocity profile is not evident in these results. These results however shows that linear theory can be used to predict the shape of radial velocity profile. However, more effort should be put into studying the amplitude. Since the predicted velocity profile is directly dependent on the density profile, it may be worth investing effort into studying how the methodology for calculating the density profile can be improved.

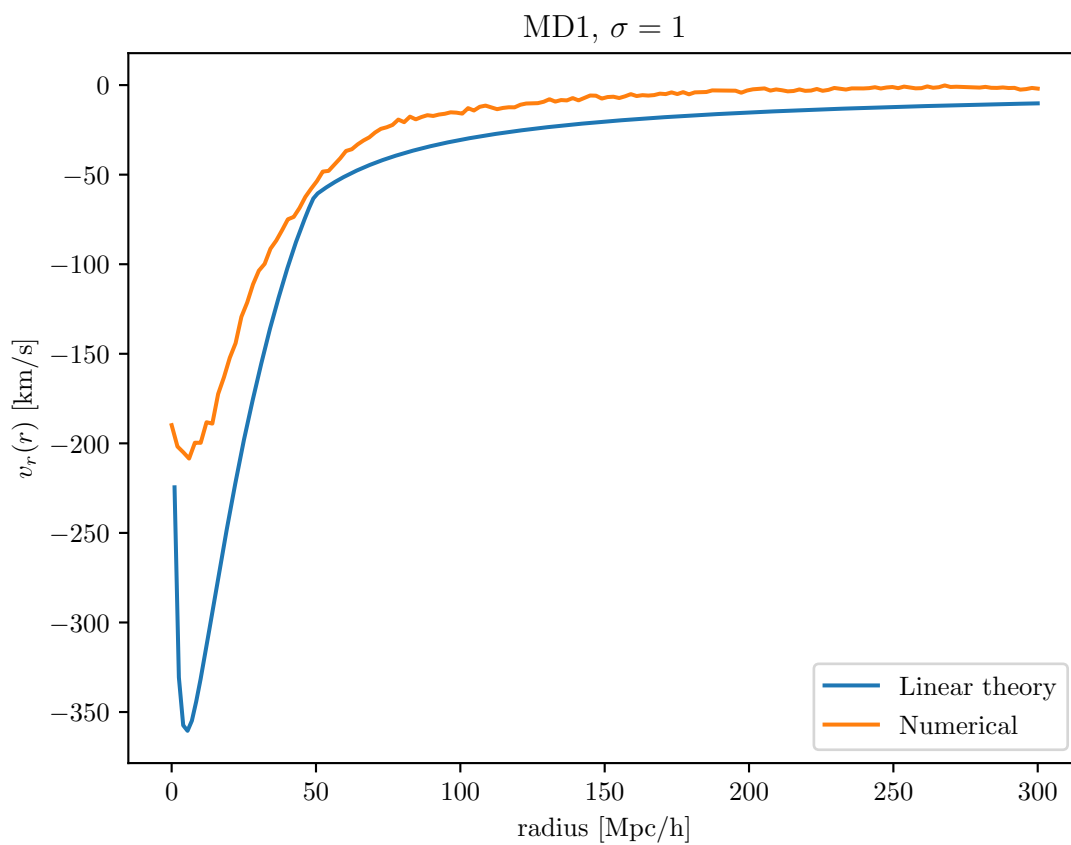


Figure 6.9: Figure showing the calculated radial velocity profile compared with the predictions from linear theory for filaments in the MD1 dataset with $\sigma = 1$.

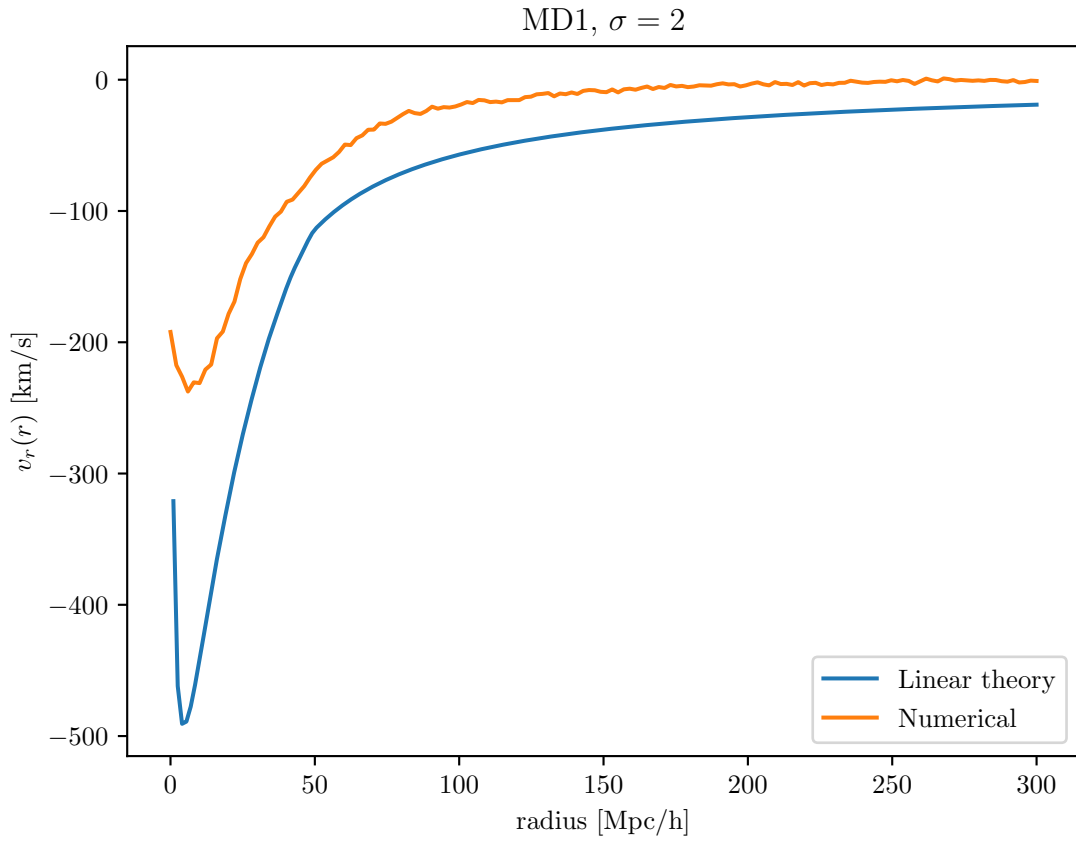


Figure 6.10: Figure showing the calculated radial velocity profile compared with the predictions from linear theory for filaments in the MD1 dataset with $\sigma = 2$.

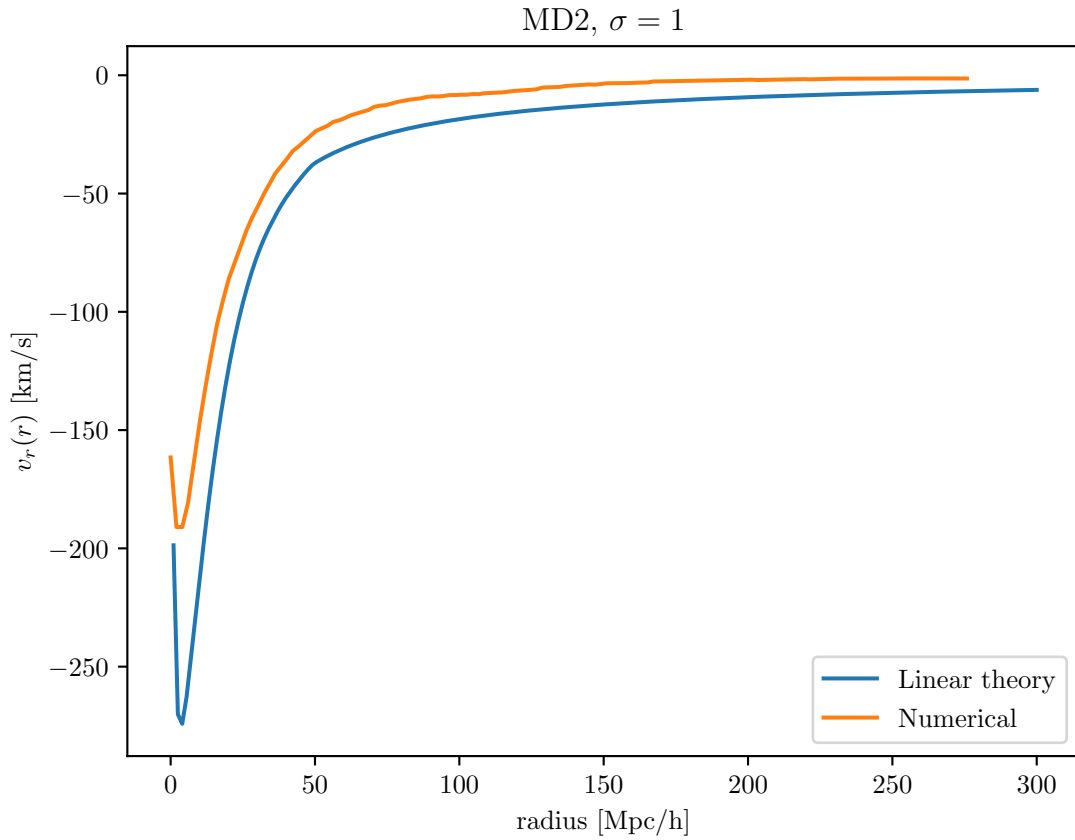


Figure 6.11: Figure showing the calculated radial velocity profile compared with the predictions from linear theory for filaments in the MD2 dataset with $\sigma = 1$.

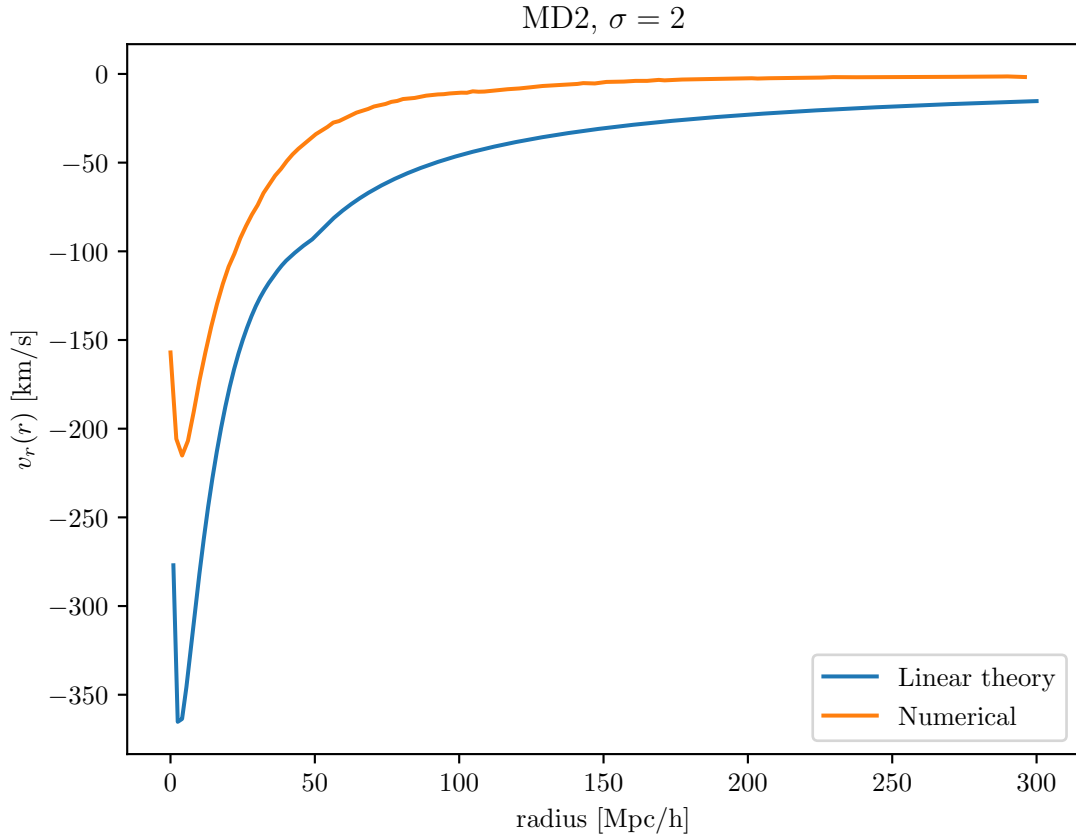


Figure 6.12: Figure showing the calculated radial velocity profile compared with the predictions from linear theory for filaments in the MD2 dataset with $\sigma = 2$.

6.2 Void analysis

6.2.1 Histograms

The effective void radius, measured in Mpc/h, is one of the quantities of voids that is derived from the void finder in the REVOLVER code. Figures 6.13, 6.13 and 6.15 shows histograms of the effective void radius found in the MD2, MD3 and MD4 datasets respectively. The histograms were calculated using radial bins with a size of 2Mpc/h. Due to a small sample size, giving lots of noise and bad fits to the cross-correlation function, the MD1 dataset was excluded from the void analysis. From the figures one can see that, for the MD2 and MD3 datasets, the bulk of the distribution of voids is in the 20Mpc/h to 80Mpc/h range. With the MD4 dataset, the distribution is shifted more towards smaller voids with the peak of the distribution at approximately

30Mpc/h. This trend can be seen throughout all three datasets tested. The more halo particles in the catalogue, the more the distribution shifts towards smaller voids. This is a reasonable result as a catalogue that is highly populated will introduce more irregularities to the density field from which the voids are defined than a more sparsely populated catalogue. The number of small voids however can introduce physical effects that may not be accurately approximated by linear theory. If the voids are small enough, non-linear interactions between halo particles may start to dominate. This can make the assumptions from section 5.1, such as the velocity of halos being purely in the radial direction, prove to not be a good approximation for their real behavior. Therefore, in order to study whether this is the case, the MD3 and MD4 datasets were split into three different cuts. One catalogue containing all voids, one catalogue containing small voids with an effective radius $20\text{Mpc/h} \leq r < 40\text{Mpc/h}$ and one catalogue with large voids containing voids with $r \geq 40\text{Mpc/h}$.

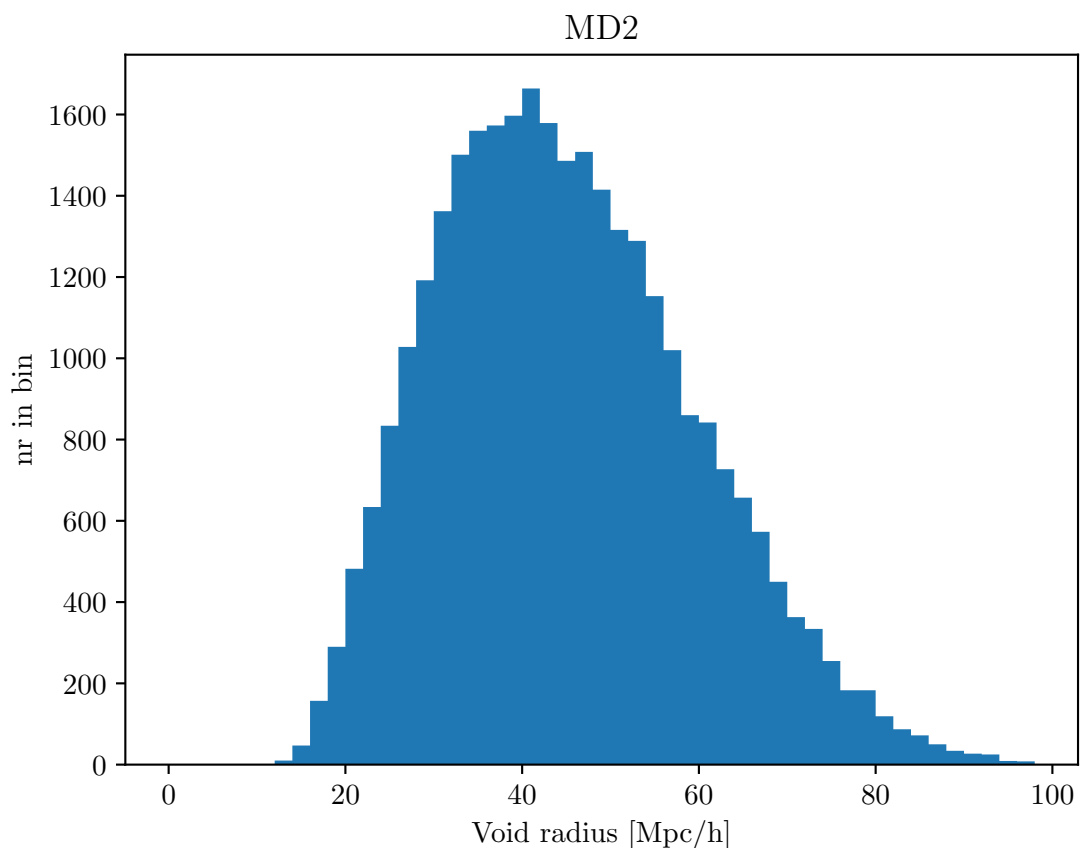


Figure 6.13: Figure showing a histogram of the effective void radius of voids found in the Multidark 2 dataset.

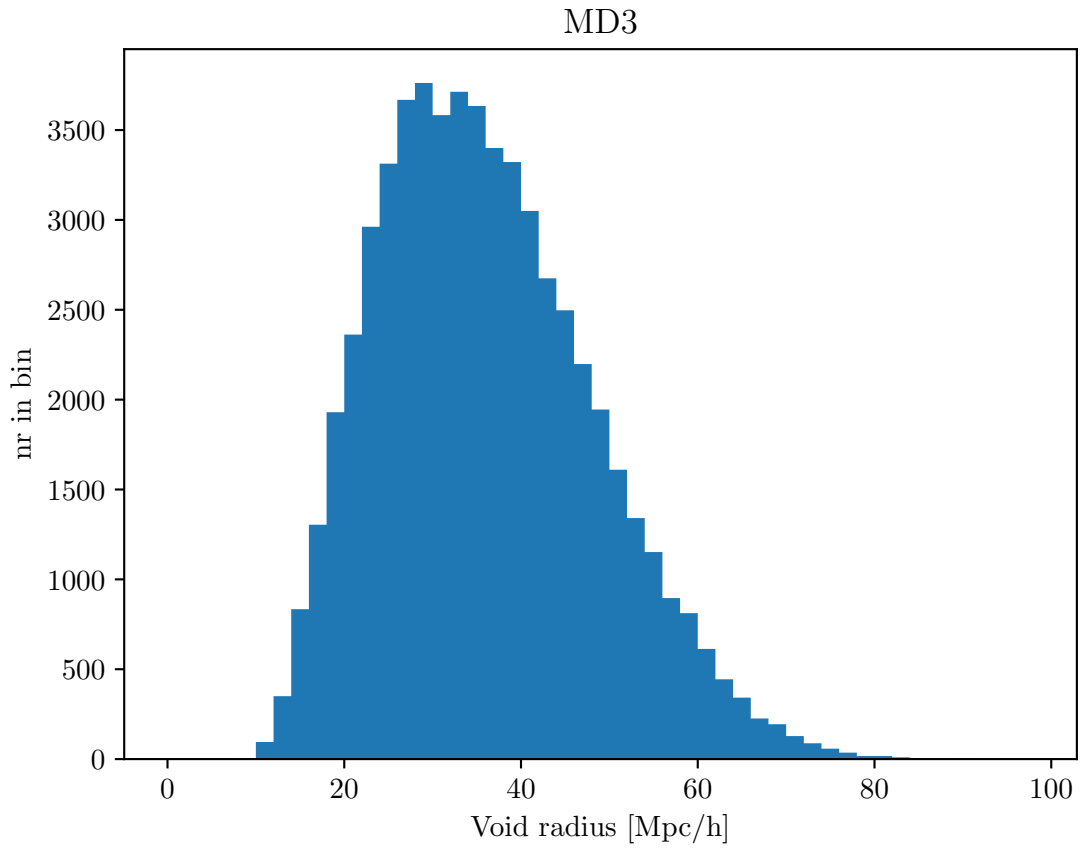


Figure 6.14: Figure showing a histogram of the effective void radius of voids found in the Multidark 3 dataset.

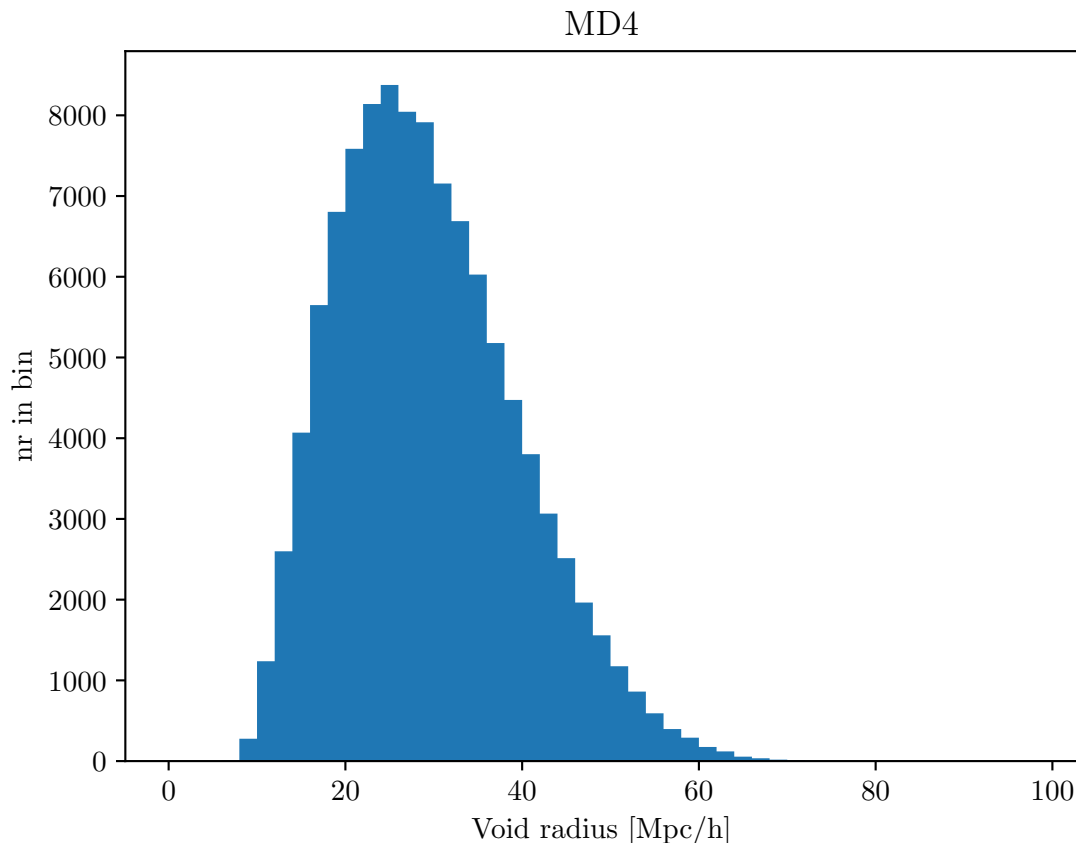


Figure 6.15: Figure showing a histogram of the effective void radius of voids found in the Multidark 4 dataset.

6.2.2 Density profiles

In accordance with the method described in section 5.3.1, the density profiles of voids $\delta(r)$ were calculated. The density profile was calculated for $2\text{Mpc}/h \leq r \leq 118\text{Mpc}/h$ and then splined using linear interpolation. The density profile for voids found in the MD2 dataset is illustrated in figure 6.16. The density profiles for the MD3 and MD4 datasets with and without cuts, are shown in figures 6.20 and 6.21 respectively. In addition to no cuts, these figures also illustrate the stacked density profile with the cuts $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ and $r \geq 40\text{Mpc}/h$. From these figures, one can see that in the center of the void the density is zero. As the radius increase, the overdensity profile starts to increase before it reaches a maximum value. This is due to the fact that overdensities are usually found on the edges of voids as matter flows away from underdensities found in the center of the void itself. After the profile reaches its maximum, the density profile

starts to decline. Far away from the void center, the average density profile converges towards the average density of the simulation volume. One can see from these figures that applying cuts will shift the peak of the overdensity. Larger voids will peak at a larger radius than the smaller voids. The smaller voids also have a more prominent overdensity than the larger voids. This can be explained by the fact that when applying the $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ cut, one observes overdensities on smaller scales. When observing overdensities on smaller scales, one averages over a smaller volume of space and fluctuations in the overdensity becomes more prominent.

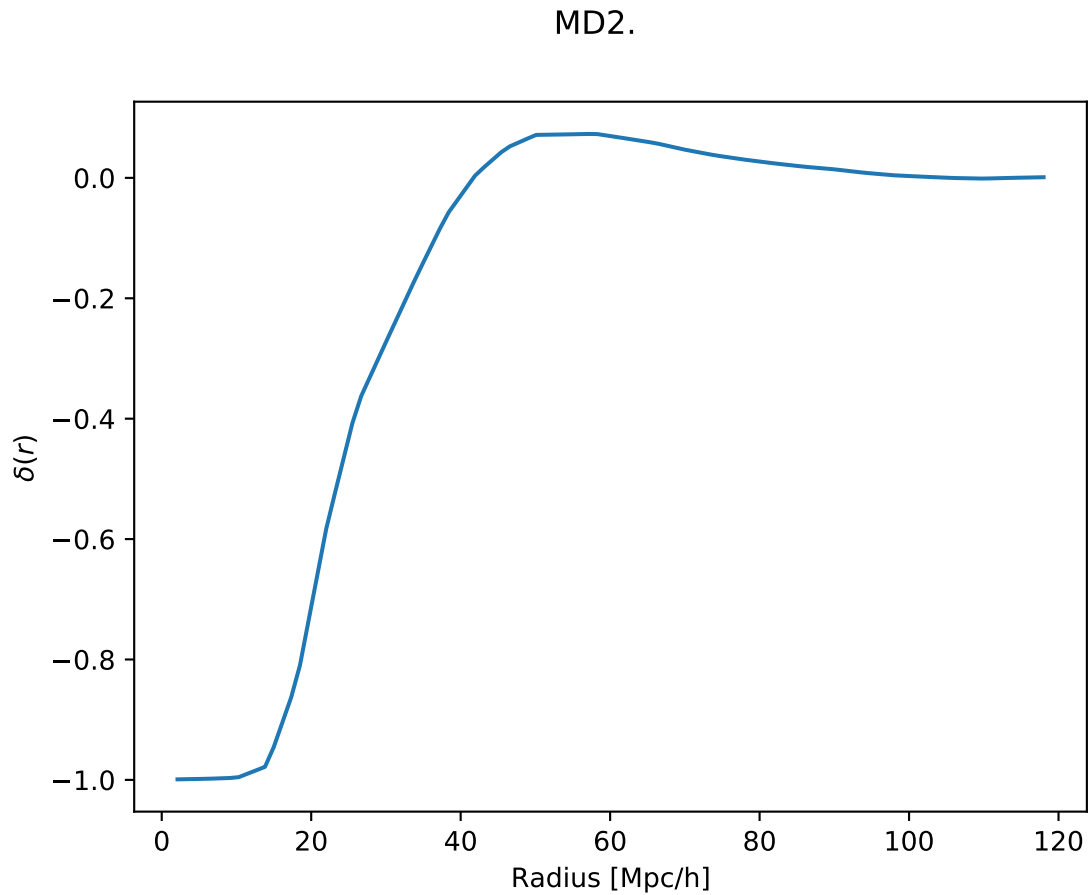


Figure 6.16: Figure showing the density profile for voids found in the Multidark 2 dataset.

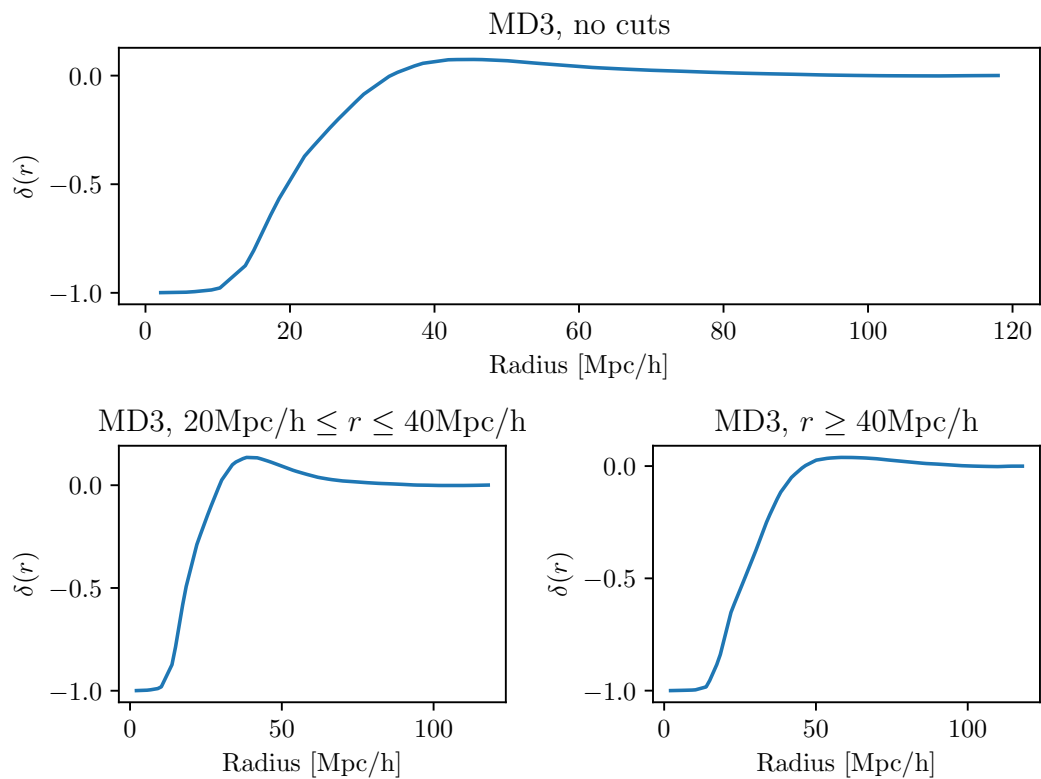


Figure 6.17: Figure showing the density profile for voids found in the Multidark 3 dataset. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ in the bottom left, and $r \geq 40\text{Mpc/h}$ in the bottom right.

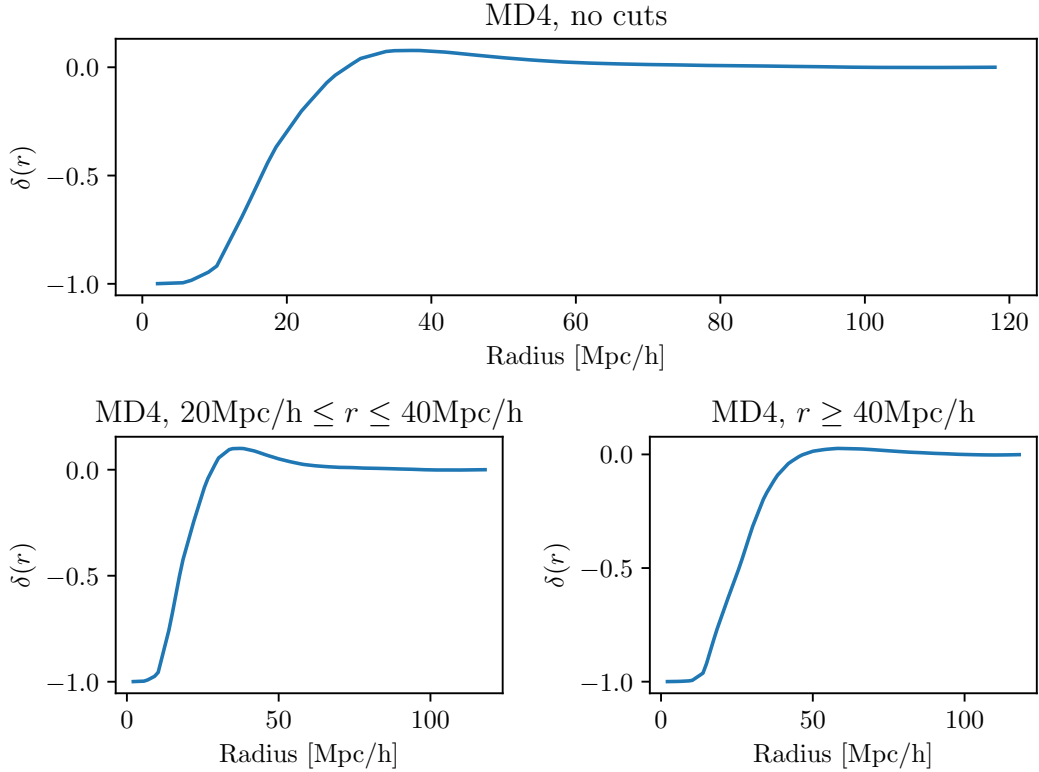


Figure 6.18: Figure showing the density profile for voids found in the Multidark 4 dataset. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ in the bottom left, and $r \geq 40\text{Mpc/h}$ in the bottom right.

These density profiles were also used to calibrate the dark matter density profile as described in section 5.6. This is shown in figures 6.19, 6.20 and 6.21 for the three datasets tested. In these figures, the blue lines labelled "Scaled $\delta_{dm}(r)$ " is the dark matter profile multiplied by a scaling factor in order for the amplitude to resemble that of $\delta_g(r)$. This factor was set to 1.85. For use in the model for the correlation function, after it was scaled with the 1.85 factor, the dark matter profile was divided by the bias term for the given MultiDark dataset it was utilized on. This is illustrated by the orange line. The bias factor for each dataset is listed in table 4.2.

The dark matter density profile is not as steep as the dark matter halo profile. This is because as $\delta_g(r)$ traces a discrete particle field, $\delta_{dm}(r)$ resembles a more fluid-like behavior. Although, in this work, the real dark matter profile for the dataset was not obtained and instead a dark matter profile from another dataset was utilized, the principle for this behavior is the same. The catalogue used to calculate $\delta_{dm}(r)$ is found from the dark matter particle field and is a point particle distribution derived from a

much larger amount of dark matter particles. The Larger amount of particles in the dark matter particle field will even out the slope of the density profile.

Dark mattter density profile MD2.

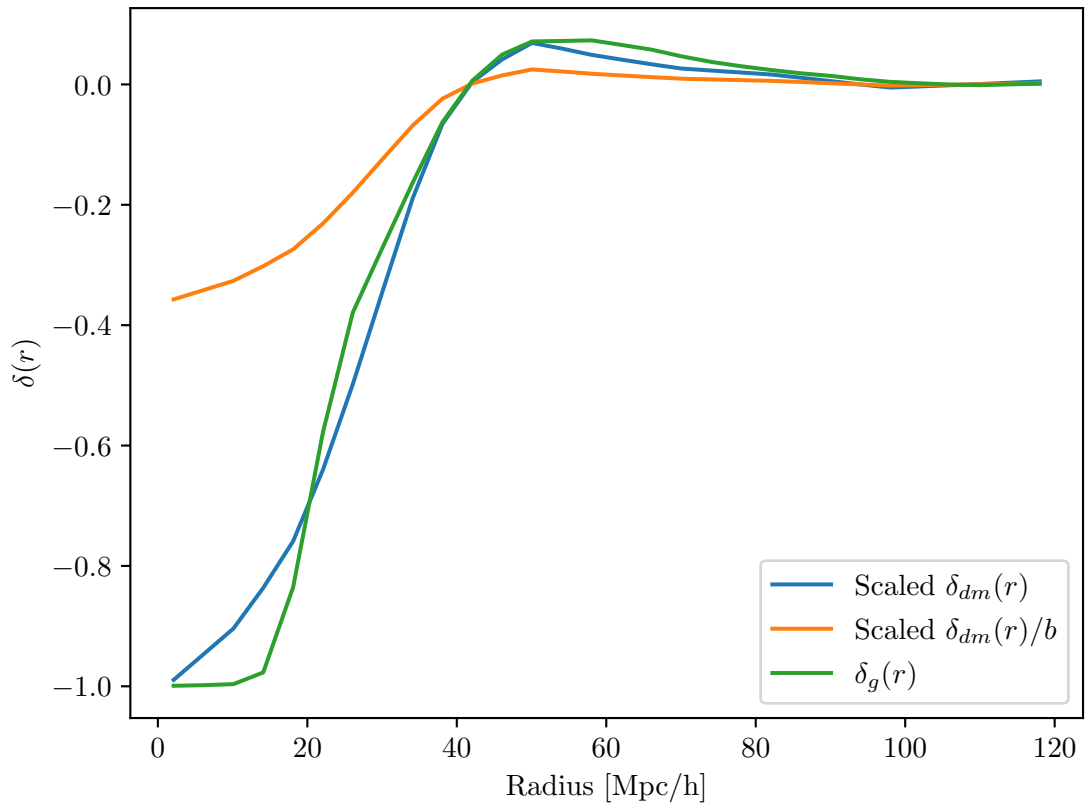


Figure 6.19: Figure showing the calibrated dark matter density profile, in accordance with the method described in section 5.6, for voids found in the Multidark 2 dataset. The blue line is the dark matter density profile multiplied by a factor 1.85 in order to make it resemble $\delta_g(r)$. The orange line is the scaled dark matter density profile divided by the bias factor for the given MultiDark dataset. The orange line is the one used in the calculation of the model.

Dark matter density profile MD3.

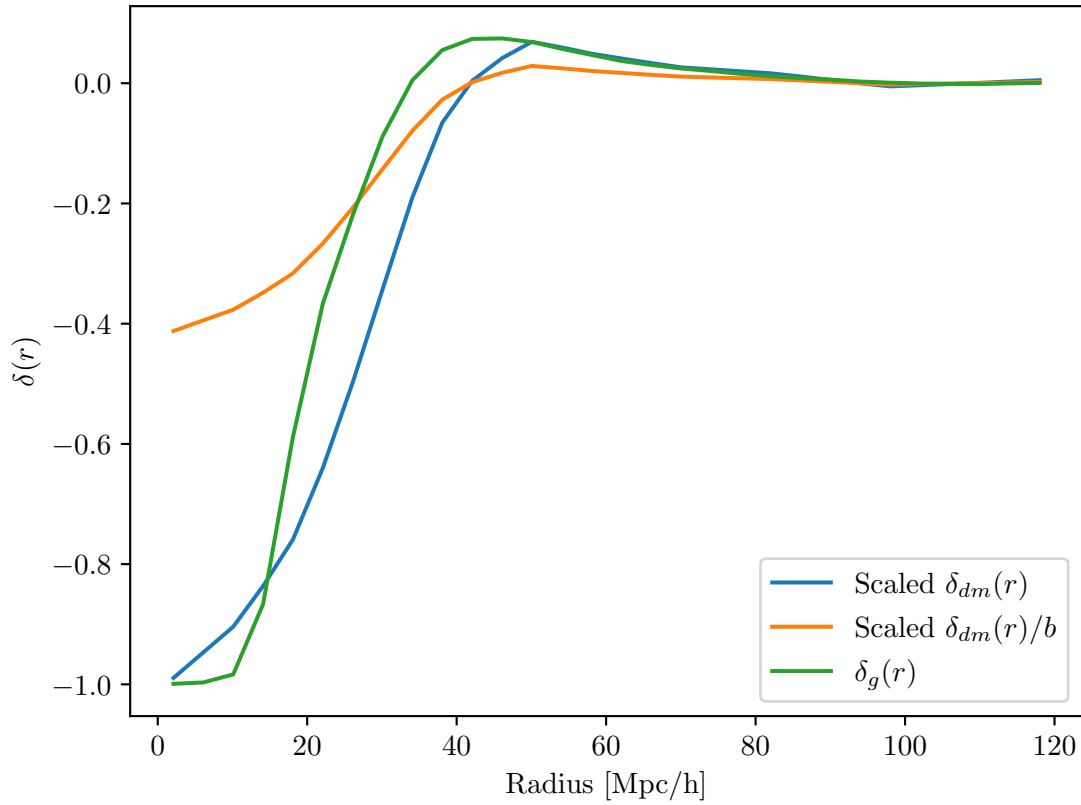


Figure 6.20: Figure showing the calibrated dark matter density profile, in accordance with the method described in section 5.6, for voids found in the Multidark 3 dataset. The blue line is the dark matter density profile multiplied by a factor 1.85 in order to make it resemble $\delta_g(r)$. The orange line is the scaled dark matter density profile divided by the bias factor for the given MultiDark dataset. The orange line is the one used in the calculation of the model.

Dark matter density profile MD4.

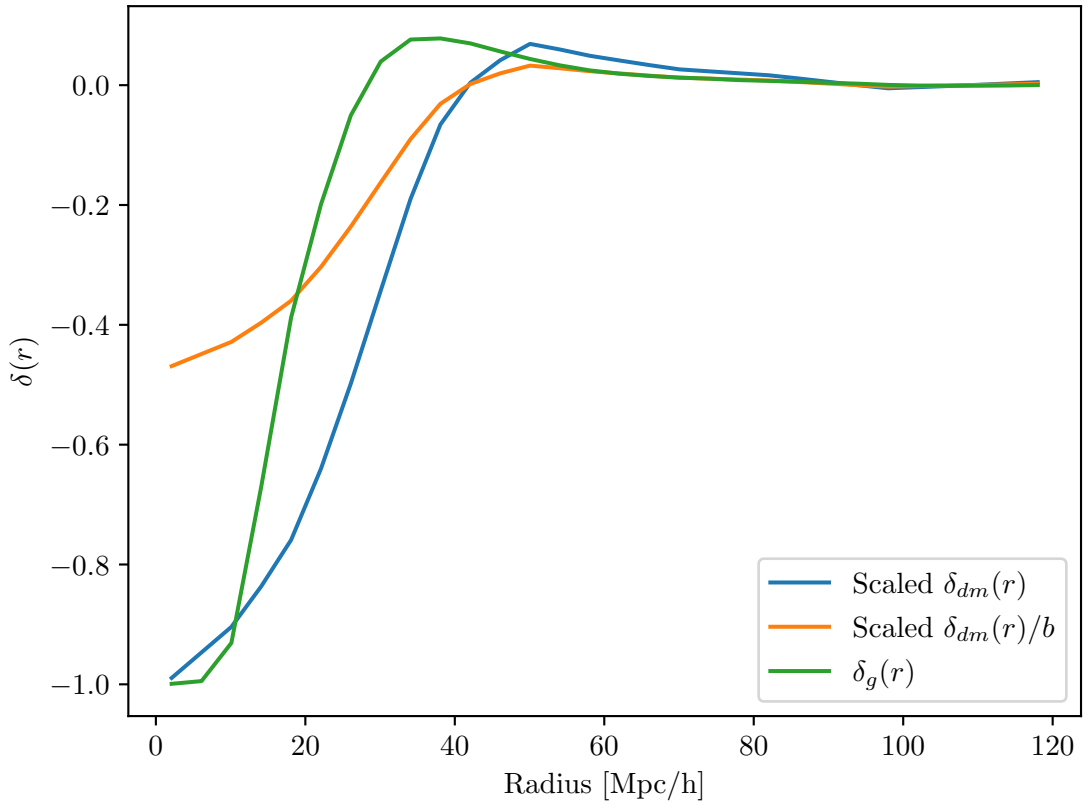


Figure 6.21: Figure showing the calibrated dark matter density profile, in accordance with the method described in section 5.6, for voids found in the Multidark 4 dataset. The blue line is the dark matter density profile multiplied by a factor 1.85 in order to make it resemble $\delta_g(r)$. The orange line is the scaled dark matter density profile divided by the bias factor for the given MultiDark dataset. The orange line is the one used in the calculation of the model.

In order to calculate the correlation function given in equations 5.32 and 5.39, the density contrast for voids $\Delta_v(r)$, given in equation 5.11, was calculated for all the datasets with the aforementioned cuts. This is illustrated in the figures 6.22, 6.23 and 6.24. This quantity shows the integrated overdensity as a function of radius.

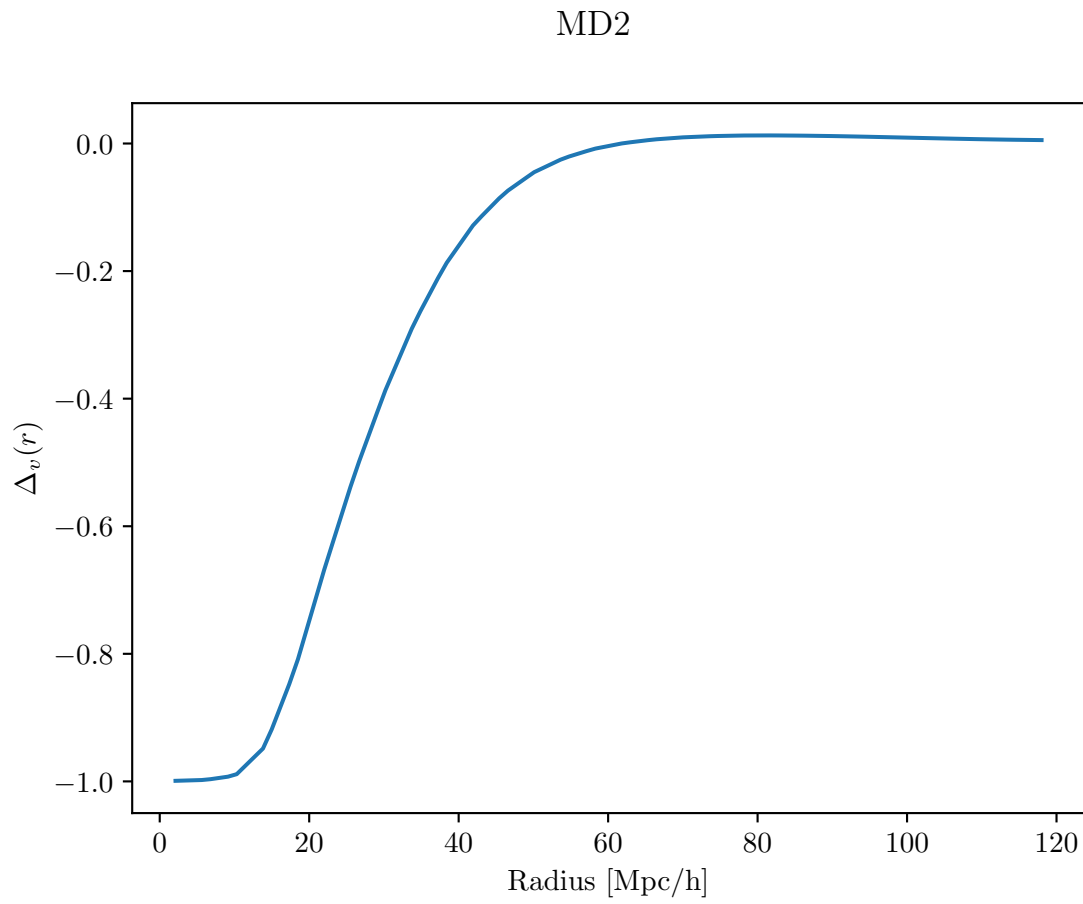


Figure 6.22: Figure showing the density contrast $\Delta_v(r)$ for voids found in the Multidark 2 dataset.

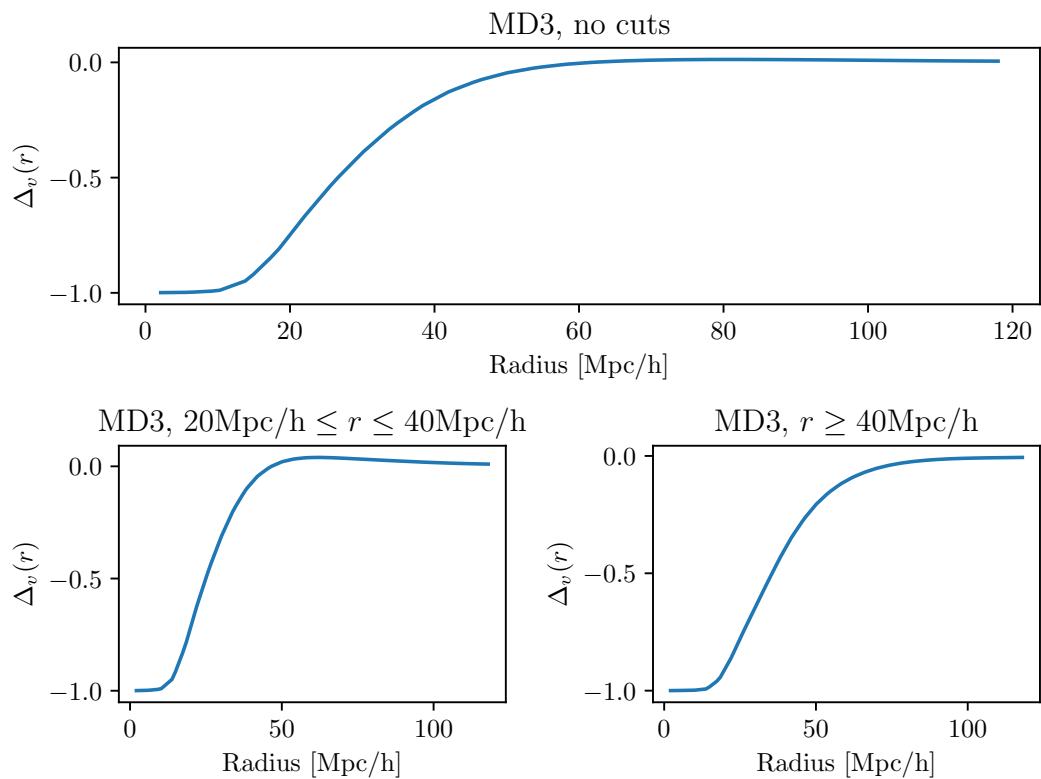


Figure 6.23: Figure showing the density contrast $\Delta_v(r)$ for voids found in the Multidark 3 dataset. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ in the bottom left, and $r \geq 40\text{Mpc}/h$ in the bottom right.

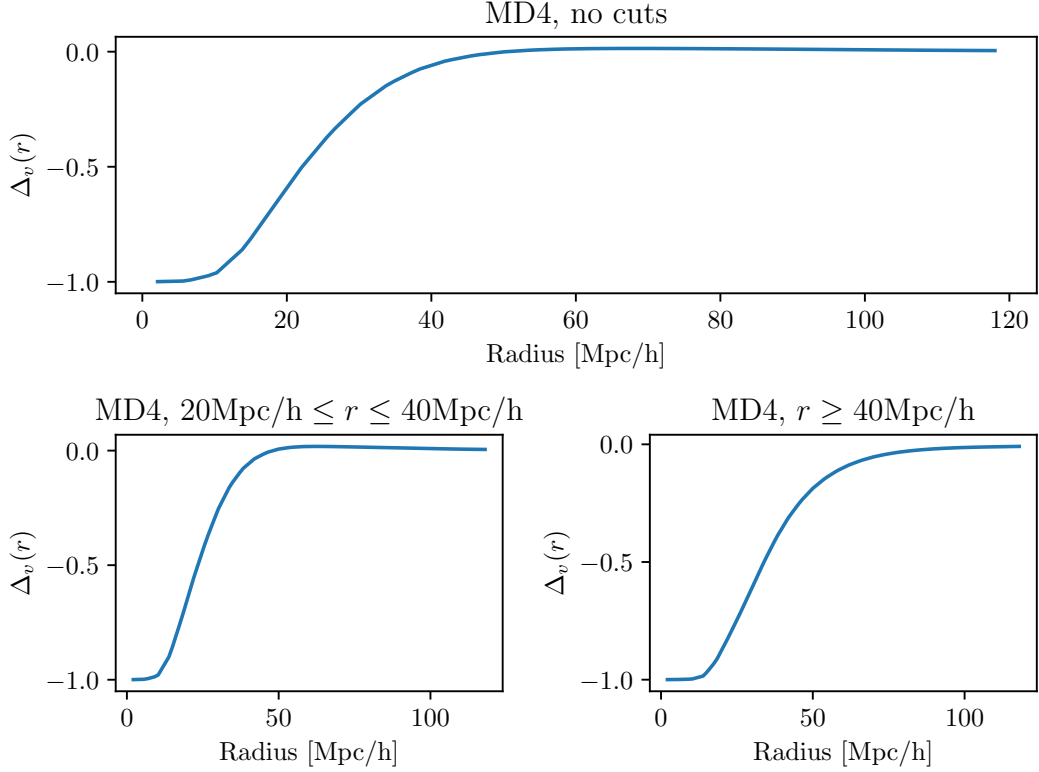


Figure 6.24: Figure showing the density contrast $\Delta_v(r)$ for voids found in the Multidark 4 dataset. Here the radius of voids considered are in the radius range of: all voids in top figure, $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ in the bottom left, and $r \geq 40\text{Mpc}/h$ in the bottom right.

6.2.3 Radial velocity profiles

The radial velocity profile $v_r(r)$ for all three datasets was calculated in accordance with the method described in section 5.3.2. The numerically calculated velocity profile was calculated for $2\text{Mpc}/h \leq r \leq 118\text{Mpc}/h$ using 30 linearly space bins. After this, the radial velocity profile $v_r(r)$ was splined using linear interpolation. These were also compared with the predicted radial velocity from linear theory, given in equation 5.12. This is illustrated in figures 6.25, 6.26 and 6.27 for the MD2, MD3 and MD4 datasets respectively. From these figures, one can see that the measured radial velocity is in the range of approximately $150\text{km}/s$. This velocity is similar to what is measured in [49] (e.g, figure 2) and [4] (e.g, figure 2). The radial velocity is zero close to the center of the void as these areas are empty of halos. The stacked radial velocity then starts to increase. Positive direction is pointing radially inwards towards the void center, and therefore, the velocity increases in the negative direction as matter is drawn outwards,

flowing towards overdensities at the edges of voids. After the radial velocity reaches its maximum, it declines and reaches zero as other dynamics than that of the void itself starts to dominate. On these radii, the radial velocity is no longer biased along the radial vector of the void making the stacked radial velocity profile zero. From the figures for the radial velocity, one can see that when applying the $r > 40\text{Mpc/h}$ cut, there is a significant increase in velocity for the stacked velocity profile. As one considers larger voids, there is a larger underdense area. Since matter is flows away from underdense regions, the larger underdensities can cause their velocities to increase.

A qualitative comparison shows that the radial velocity predicted by linear theory matches the numerically calculated reasonably well on scales approximately $r > 30\text{Mpc/h}$. However, on smaller radii, it is evident that the predicted velocity does not capture the behavior of the numerically calculated velocity. The numerically calculated velocity has a zero gradient on small scales closer to the void center. However, the velocity predicted by linear theory starts with a steep gradient. For the datasets without cuts, one can see that this will result in the peak of the stacked velocity occurring earlier than that of the numerically calculated velocity. However, when the $r > 40\text{Mpc/h}$ cut is applied for the MD3 and MD4 datasets, both the amplitude and peak of the curve predicted by linear theory coincide with the numerically calculated a lot closer. After the velocity starts to decline, for the $r > 40\text{Mpc/h}$ cut, the numerical and predicted curve match each other in an almost identical fashion. When no cuts are applied, and for the $20\text{Mpc/h} < r < 40\text{Mpc/h}$ cuts, there is a slight offset between the two curves. This suggests that linear theory is a good approximation to the radial velocity for voids when cuts are applied and small voids subject to non-linear dynamics are left out of the catalogue.

The predicted radial velocity with the correction term, given in equation 5.13, was also calculated. For the MD2, and the MD3 and MD4 datasets without cuts, this model has an amplitude that is more in line with the numerical calculation. This is in agreement with what the author of [4] found when implementing this model. On lower radii, before the stacked velocity reaches its maximum, the model with the correction term seems to be in better agreement with the data found from the catalogue. However, the author found this model to perform better than the regular velocity proposed by linear theory on all scales. This is in contrast to what is suggested from the results found here for these datasets. On larger scales, when the numerical curve decreases, the regular predictions from linear theory seem to be in better agreement. It is worth noting that for all three samples without cuts and the samples with the $20\text{Mpc/h} < r < 40\text{Mpc/h}$ cuts, there is a discontinuity in the velocity predicted by the model with the correction term. This discontinuity stems from the Γ factor in equation 5.13. When $\Delta_v(r)$ changes from negative to positive, the Γ factor changes from -1 to $+1$. This gives rise to the discontinuity observed in the velocity predicted by this model. This discontinuity makes the model appear better than the regular linear theory model on larger radii. For the MD3 and MD4 datasets with the $r > 40\text{Mpc/h}$ cuts applied, the amplitude predicted by the regular model is more in line with the numerical data. On these datasets, with the given cuts, the model with the correction term underestimates the peak of the velocity. On larger radii, there is good compliance between the numerically calculated model and

the velocity predicted by linear theory using these cuts. For the model with the correction term, however, there is a small discrepancy on these scales. Overall, these results suggest that the model predicted by linear theory is a better predictor for the velocity of halos around voids when cuts excluding small voids are applied. The model with the correction term, on the other hand, seem to model the smaller voids better. This exemplified in the amplitude being more in line with numerical data. It is therefore not clear if there is a significant difference between the models with these cuts when doing a purely qualitative comparison.

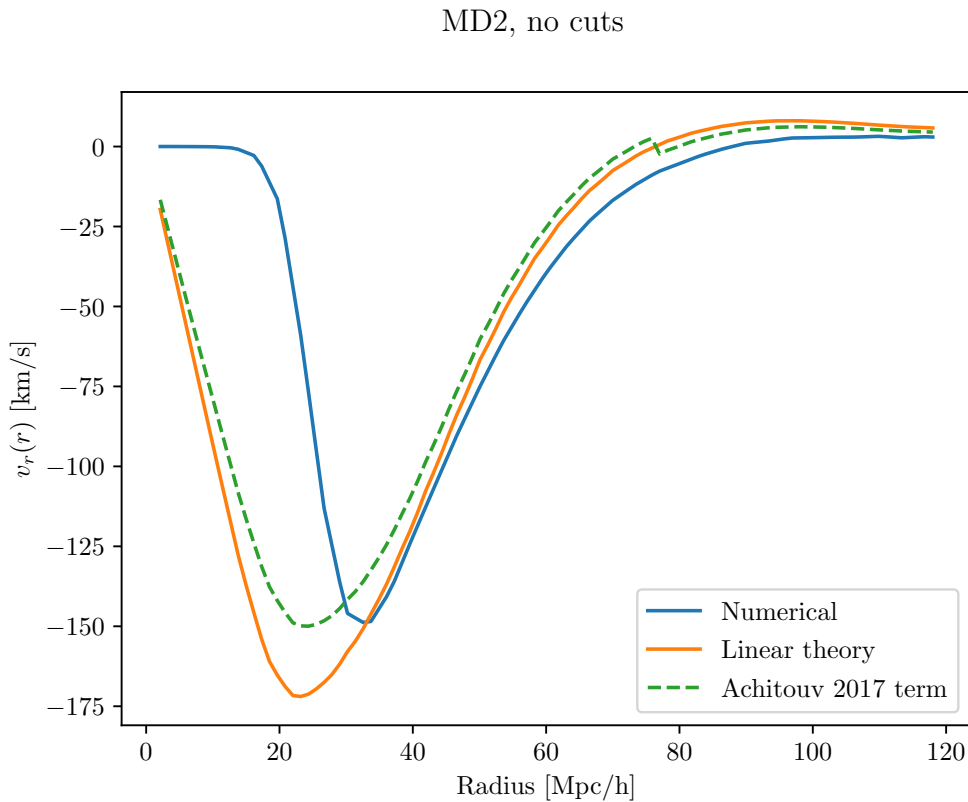


Figure 6.25: Figure showing the radial velocity profile of halos around voids found in the Multidark 2 dataset. Here it is compared with the prediction from linear theory, given in equation 5.12, as well as the model with the correction term, given in equation 5.13.

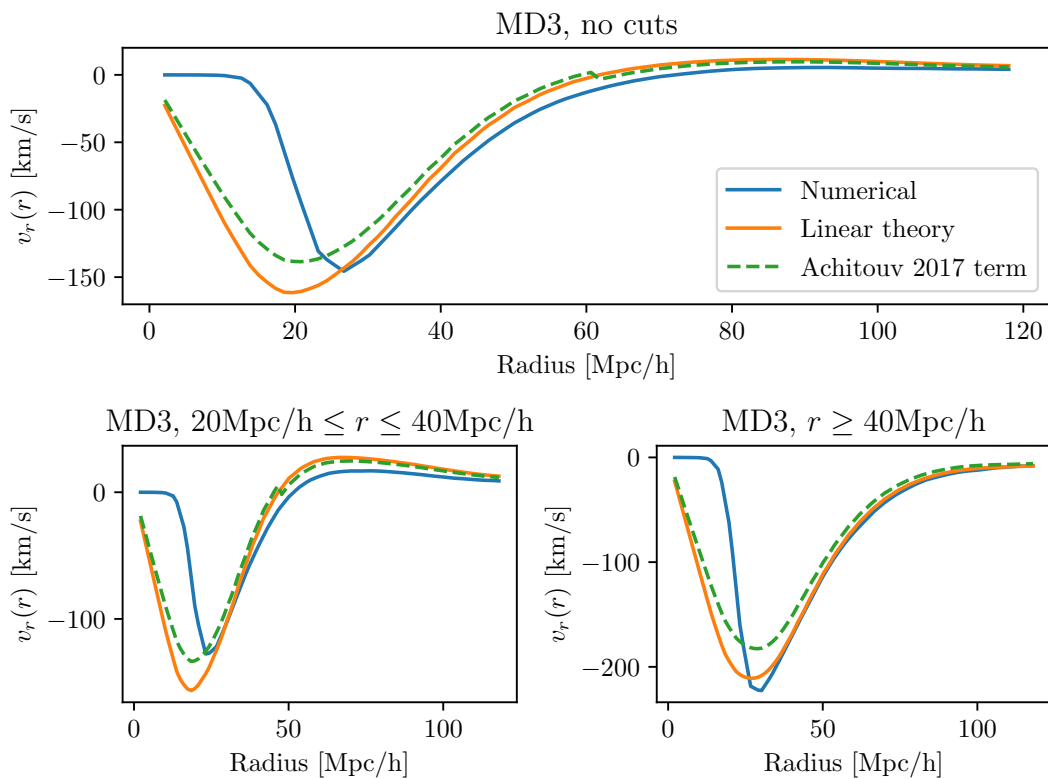


Figure 6.26: Figure showing the radial velocity profile of halos around voids found in the Multidark 3 dataset. Here it is compared with the prediction from linear theory, given in equation 5.12, as well as the model with the correction term, given in equation 5.13. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ in the bottom left, and $r \geq 40\text{Mpc/h}$ in the bottom right.

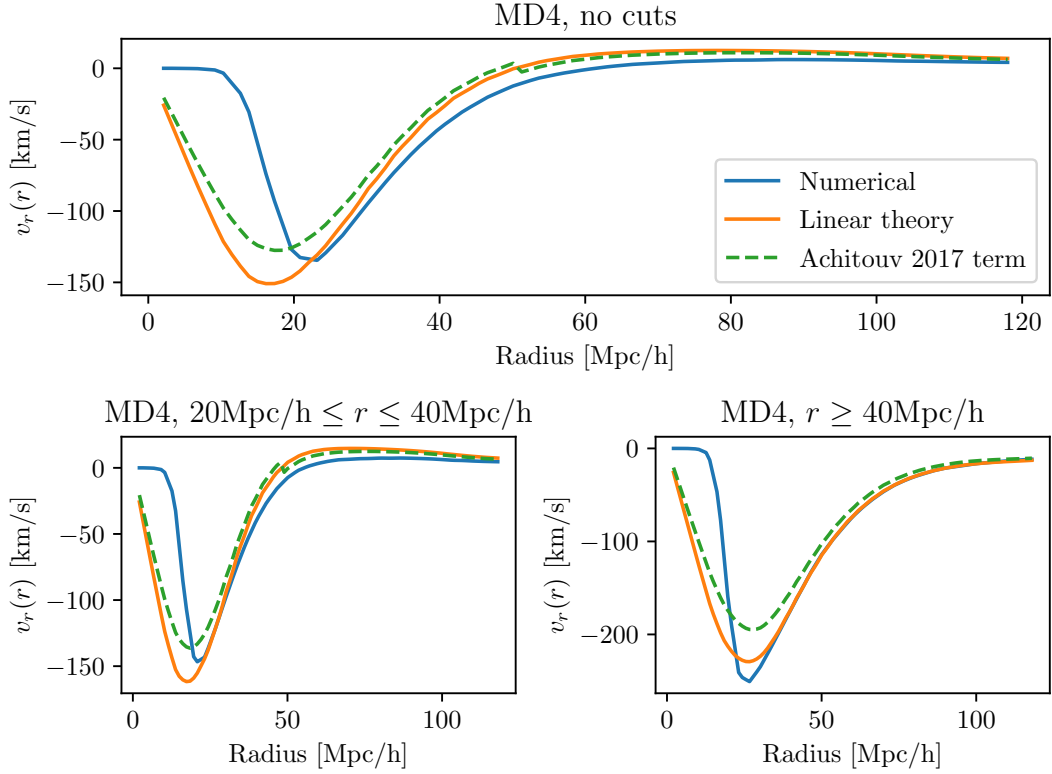


Figure 6.27: Figure showing the radial velocity profile halos around voids found in the Multidark 4 dataset. Here it is compared with the prediction from linear theory, given in equation 5.12, as well as the model with the correction term, given in equation 5.13. Here the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ in the bottom left, and $r \geq 40\text{Mpc/h}$ in the bottom right.

The last major component entering the correlation function in equation 5.39 is the velocity dispersion σ_{v_z} . The velocity dispersion was calculated for $2\text{Mpc/h} \leq r \leq 118\text{Mpc/h}$ using 30 linearly space bins and was then splined using linear interpolation. This quantity was calculated for all three datasets with and without the aforementioned cuts and is illustrated in figures 6.28, 6.29 and 6.30 for the three datasets respectively. As in [17] and [4], the velocity dispersion profile converges to a value of around 300–350 km/s, giving good agreement with previous works. However, close to the void center the dispersion profile is noisy. Close to the void centers there are on average very few dark matter halos. When there are few halos that one can utilize to calculate the variance, one also gets a low sample size which will introduce noise when calculating the dispersion of the velocity. This may give unreliable results later on when looking at small scales. However, on small scales the correlation function is practically zero and one can argue that this should not have a profound effect on the results. For future

improvements, one could manually set the values at small scales to a fixed value to neglect this problem.

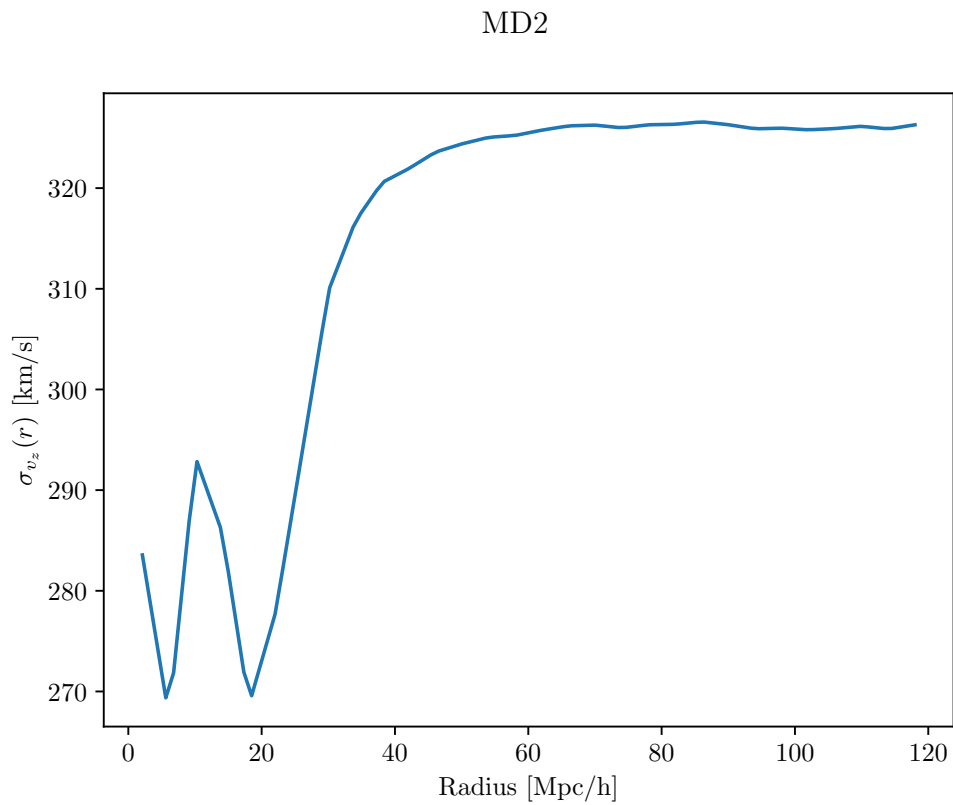


Figure 6.28: Figure showing the velocity dispersion, given by equation 5.23, for halos around voids found in the Multidark 2 dataset.

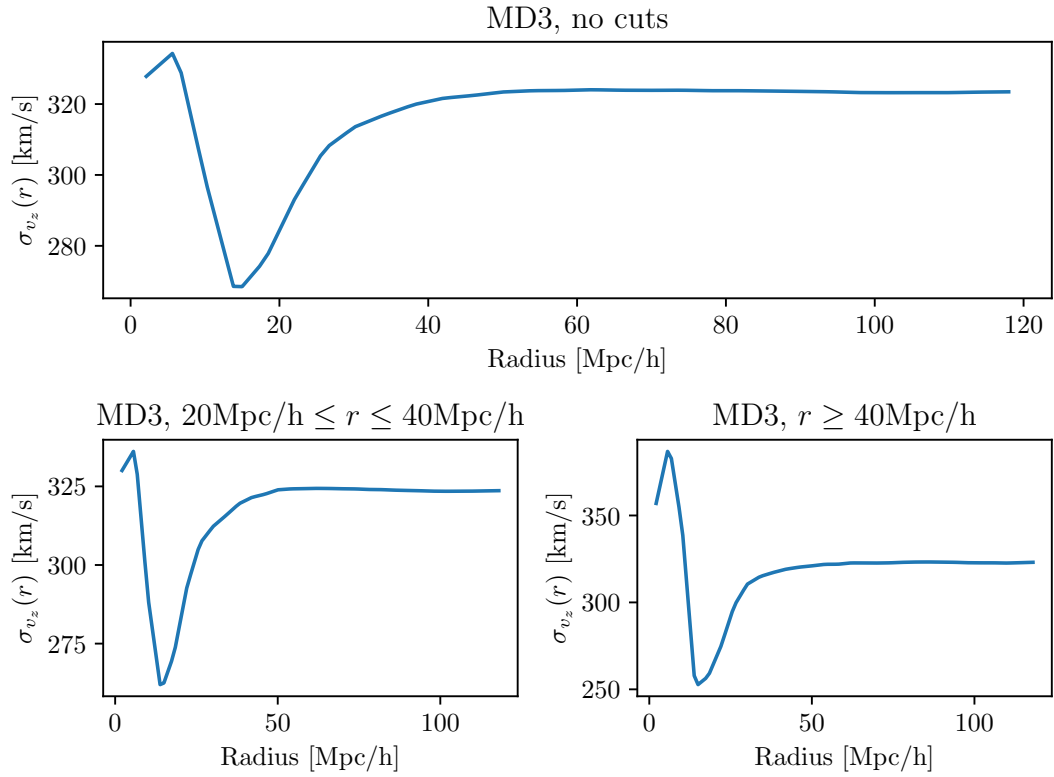


Figure 6.29: Figure showing the velocity dispersion, given by equation 5.23, for halos around voids found in the Multidark 3 dataset. In this figure, the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ in the bottom left, and $r \geq 40\text{Mpc/h}$ in the bottom right.

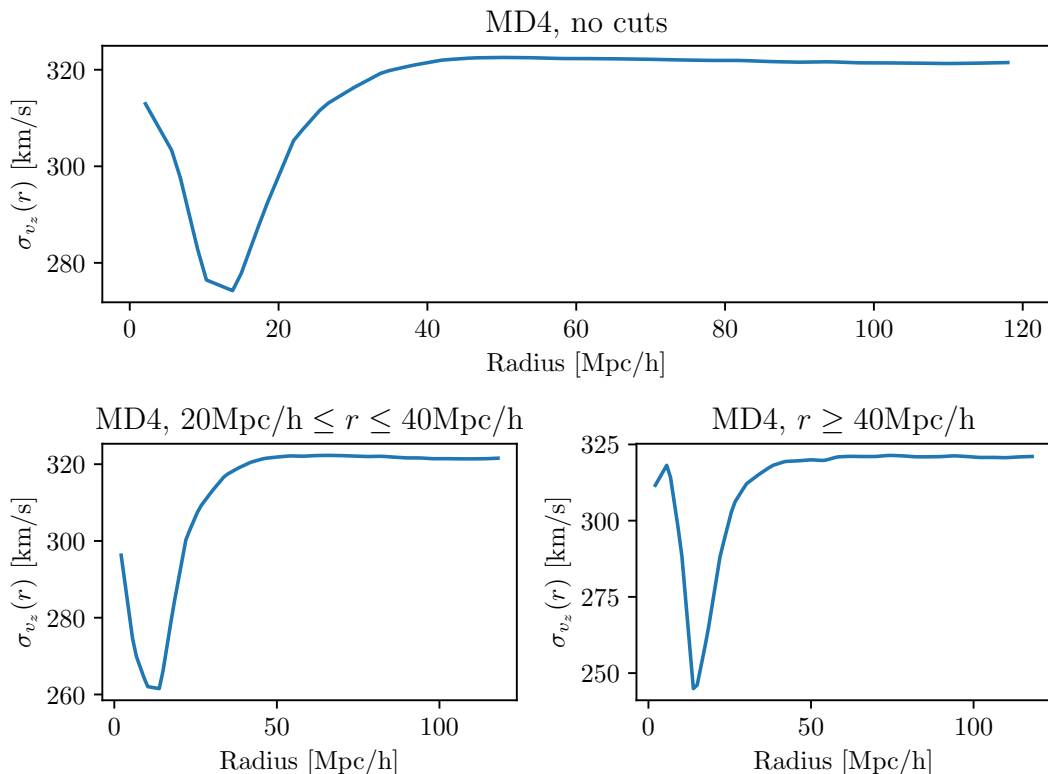


Figure 6.30: Figure showing the velocity dispersion, given by equation 5.23, for halos around voids found in the Multidark 4 dataset. In this figure, the radius of voids considered are in the radius range of: all voids in the top figure, $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ in the bottom left, and $r \geq 40\text{Mpc/h}$ in the bottom right.

6.3 Parameter fits

When fitting the parameters to data, there is especially one parameter that is interesting. This parameter is ϵ . Through the relations given in equations 5.52 and 5.51 this parameter is directly linked to the cosmology. The parameters β , when using a linear bias approximation, and $f\sigma_8$, when using a dark matter profile, are also of interest. In this case, since I do not possess the dark matter profile for our dataset and instead try to adapt a dark matter profile from the HOD mock galaxy catalogue, there is limited information that can be read from $f\sigma_8$ in this case. σ_v and r_{scale} are treated as nuisance parameters.

6.3.1 Linear bias approximation

The model using a linear bias approximation and the velocity derived from linear theory was compared with data through a maximum likelihood analysis, as described in section 5.7. For all the three datasets, with cuts, the maximum likelihood value and the 68% confidence interval are shown in table 6.1. Figure 6.31 shows the best fits for the parameters ϵ , β and σ_v for the MD2 dataset with no cuts. All the likelihood estimate plots have the fiducial values from the dataset overplotted as lines and the maximum likelihood values as a cross. Here one can see that for the MD2 dataset without cuts, the fiducial values for β and ϵ lies within the 68% confidence interval. The fits to σ_v have the fiducial value outside of the 68% confidence interval. Since this is treated as a nuisance parameter, much attention will not be directed at studying or discussing the fits to σ_v . As one can see from the histogram over effective void radii found in the dataset, shown in figure 6.13, this dataset had the distribution of voids shifted the most towards higher radii. As the peak of the distribution is located at approximately 40Mpc/h, it seems likely that the majority of voids found in the dataset are large enough so that the approximations from linear theory are sufficient to model the physical behavior of dark matter halos around voids in this dataset.

The same analysis was applied to the MD3 dataset without cuts. This is shown in figure 6.32. Here one can clearly see that the model provides bad fits for β . Although the model performs well when providing fits for ϵ , the β parameter is also important for studying the underlying cosmology. As one can see from the histogram of the effective radii for voids in the MD3 dataset, shown in figure 6.14, one can see that, in comparison with the histogram for voids in the MD2 dataset, the distribution is shifted dramatically towards smaller voids. When including all voids in the analysis, a large number of voids may then include dark matter halos subject to effects not properly approximated by linear theory thus resulting in a worse performance. Figure 6.40 shows the same analysis but here only voids with effective radius $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$ are considered. Here one can see that although the model provides good fits for ϵ , the model struggles with fitting β for this dataset. In figure 6.41 the model was tested on the MD3 dataset but only considering voids with effective radius $r \geq 40\text{Mpc/h}$. Here one can see that with the small voids excluded, the model provides good fits for all the parameters. The maximum likelihood values for β and ϵ are very close to their fiducial values. This strengthens the suspicion that non-linear effects found in small voids make the model perform worse. Thus when using a dataset with a large number of particles, which may result in a large number of small voids, applying cuts may prove crucial for gaining the correct information about the parameters if one were to apply these methods on observational data where the fiducial values are unknown.

From the histogram of effective void radii for MD4 dataset, shown in figure 6.15, it is evident that this dataset, which includes the most particles, has the distribution of effective radii shifted the most towards small voids. This means that a large fraction of the voids in this dataset may be subject to non-linear effects. Figure 6.35 shows the model applied to the MD4 dataset without cuts. Here one can see that although decent fits are provided to σ_v and ϵ , the performance of the model when fitting β is even

Dataset	ϵ	β	σ_v [km/s]
MD2	0.9956 ± 0.0093	0.2947 ± 0.0317	392.789 ± 33.298
MD3 no cuts.	0.9982 ± 0.0089	0.3218 ± 0.0325	377.544 ± 27.340
MD3 $20\text{Mpc/h} \leq r \leq 40\text{ Mpc/h}$	1.0002 ± 0.0074	0.3232 ± 0.0290	356.577 ± 24.738
MD3 $r \geq 40\text{Mpc/h}$	0.99904 ± 0.01002	0.3678 ± 0.0335	308.923 ± 47.966
MD4	1.00354 ± 0.00878	0.3050 ± 0.0339	327.840 ± 24.055
MD4 $20\text{Mpc/h} \leq r \leq 40\text{ Mpc/h}$	1.0004 ± 0.00078	0.36657 ± 0.03093	308.318 ± 23.836
MD4 $r \geq 40\text{ Mpc/h}$	0.9981 ± 0.0104	0.4097 ± 0.0340	256.351 ± 55.582

Table 6.1: Best fit values for the parameters ϵ , β and σ_v for the different datasets using a linear bias approximation.

worse than for the MD3 dataset without cuts. Figure 6.36 shows the model applied on the MD4 dataset only considering voids with effective radius $20\text{Mpc/h} \leq r \leq 40\text{Mpc/h}$. Again, one can see that the model is not sufficient to accurately fit β when voids of this size are considered. With this cut, the model performs slightly better than with no cuts. This may be due to very small voids with radius $r < 20\text{Mpc/h}$ not being considered. When applying cuts, including only voids with $r \geq 40\text{Mpc/h}$, as shown in figure 6.35, one can see that the model is accurately picking out values close to their fiducial values for the parameters β and ϵ . These results indicate that linear theory may not be sufficient to accurately model small voids with radii in the 20Mpc/h range. Future efforts could be put into better understanding at what sizes of voids linear theory breaks down. This is in order to maximize the amount of void-halo pairs and gain better statistics for extracting the information from the density field.

MD2 with linear bias approximation.

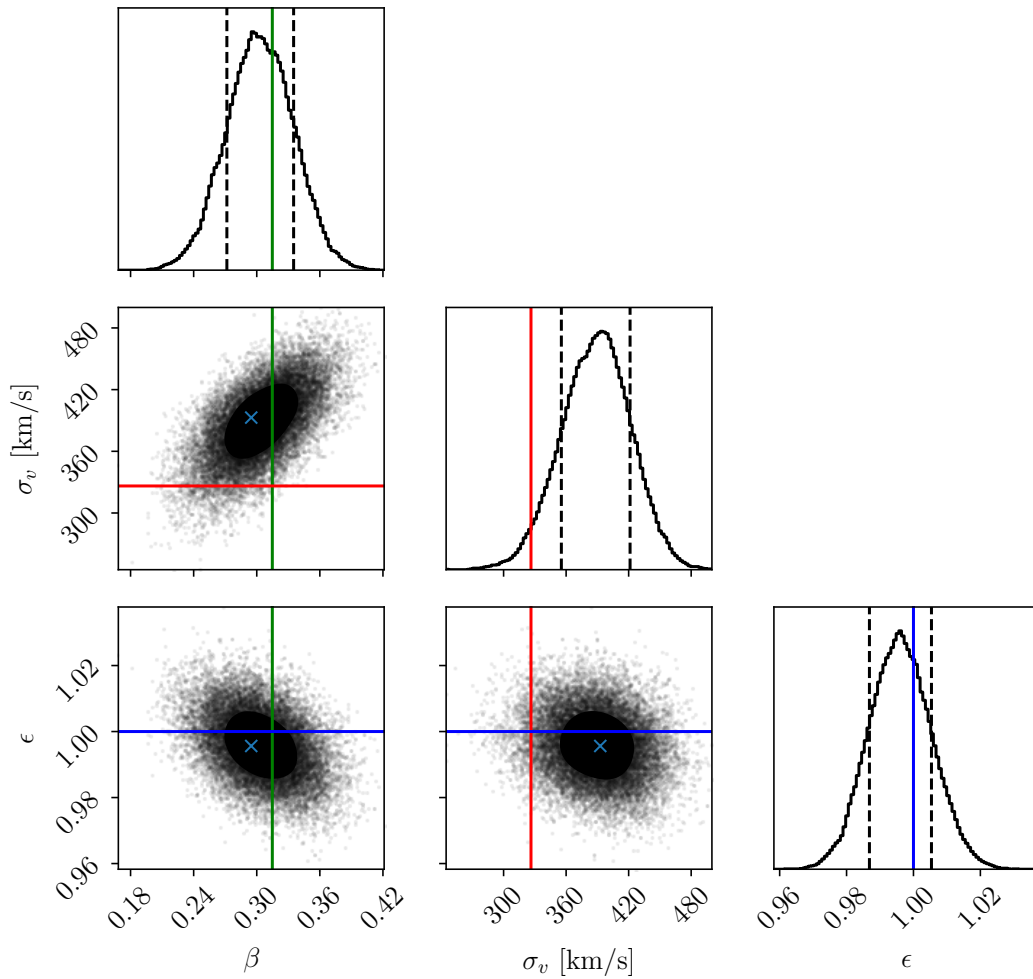


Figure 6.31: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 2 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. The fiducial value for each parameter is overplotted.

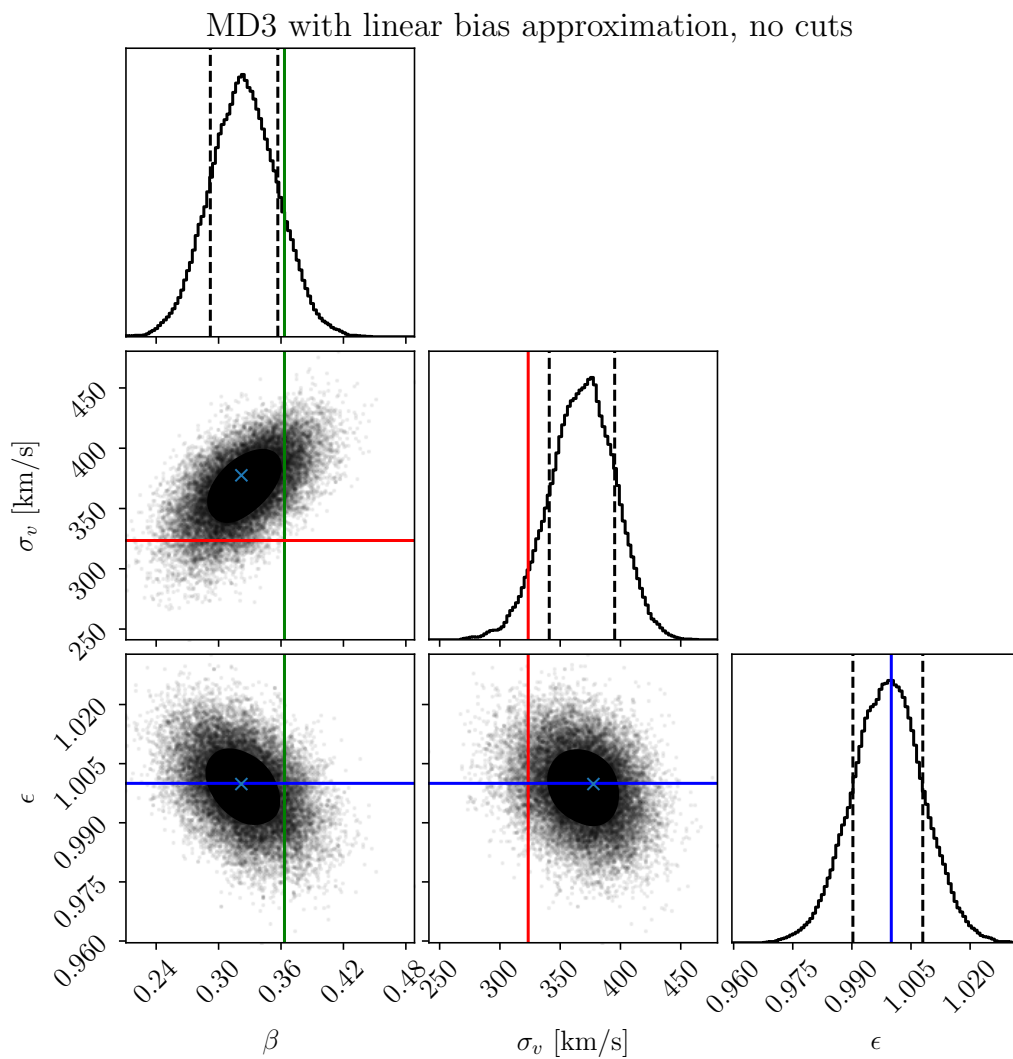


Figure 6.32: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. The fiducial value for each parameter is overplotted.

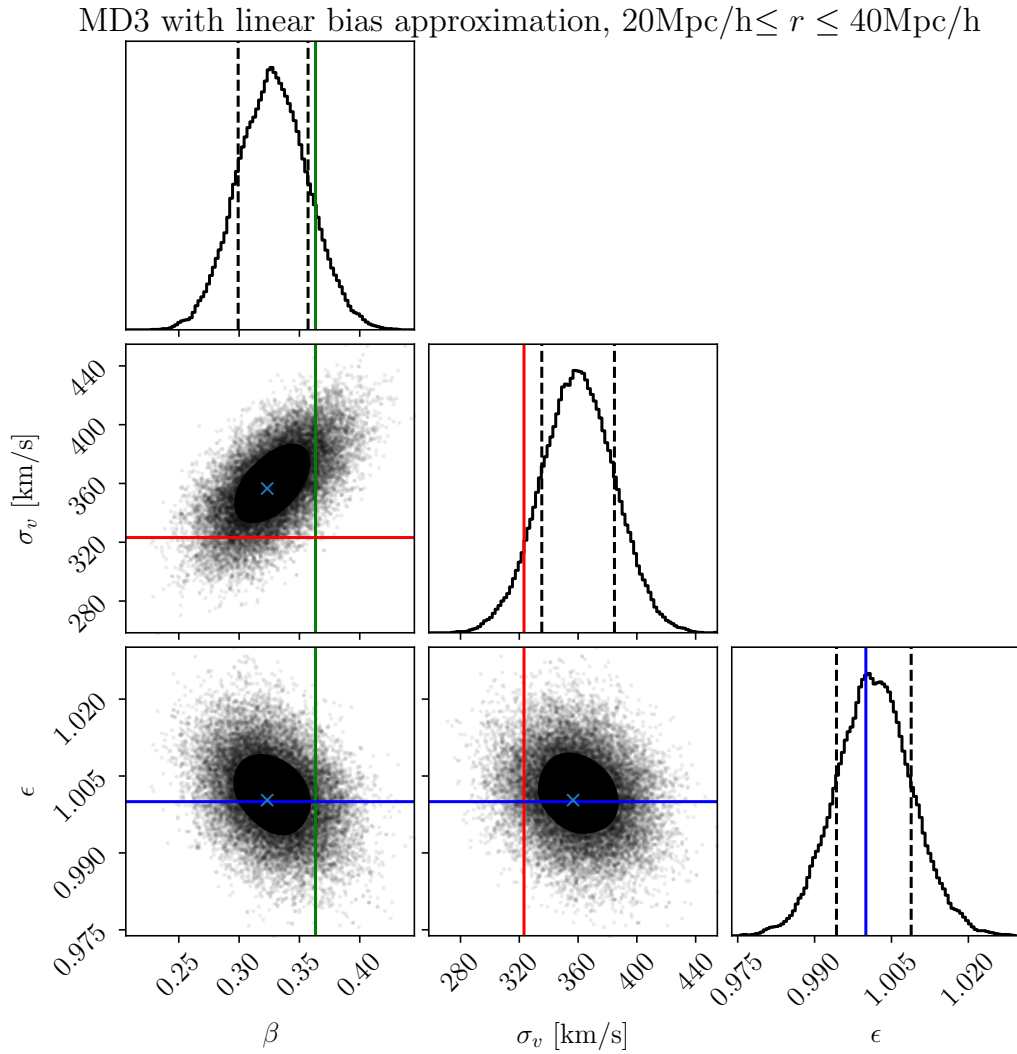


Figure 6.33: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. In this figure, only voids with effective radius $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ are considered. The fiducial value for each parameter is overplotted.

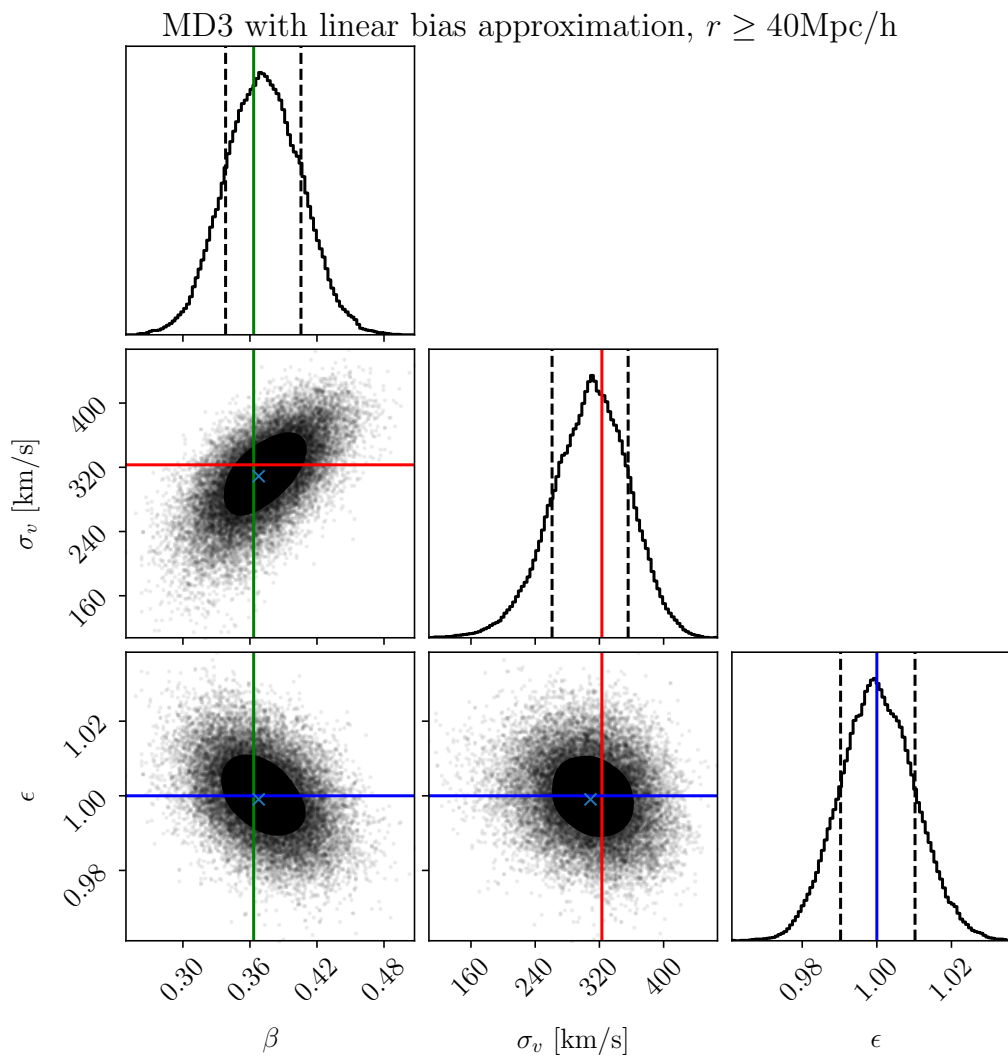


Figure 6.34: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. In this figure, only voids with effective radius $\geq 40\text{Mpc}/h$ are considered. The fiducial value for each parameter is overplotted.

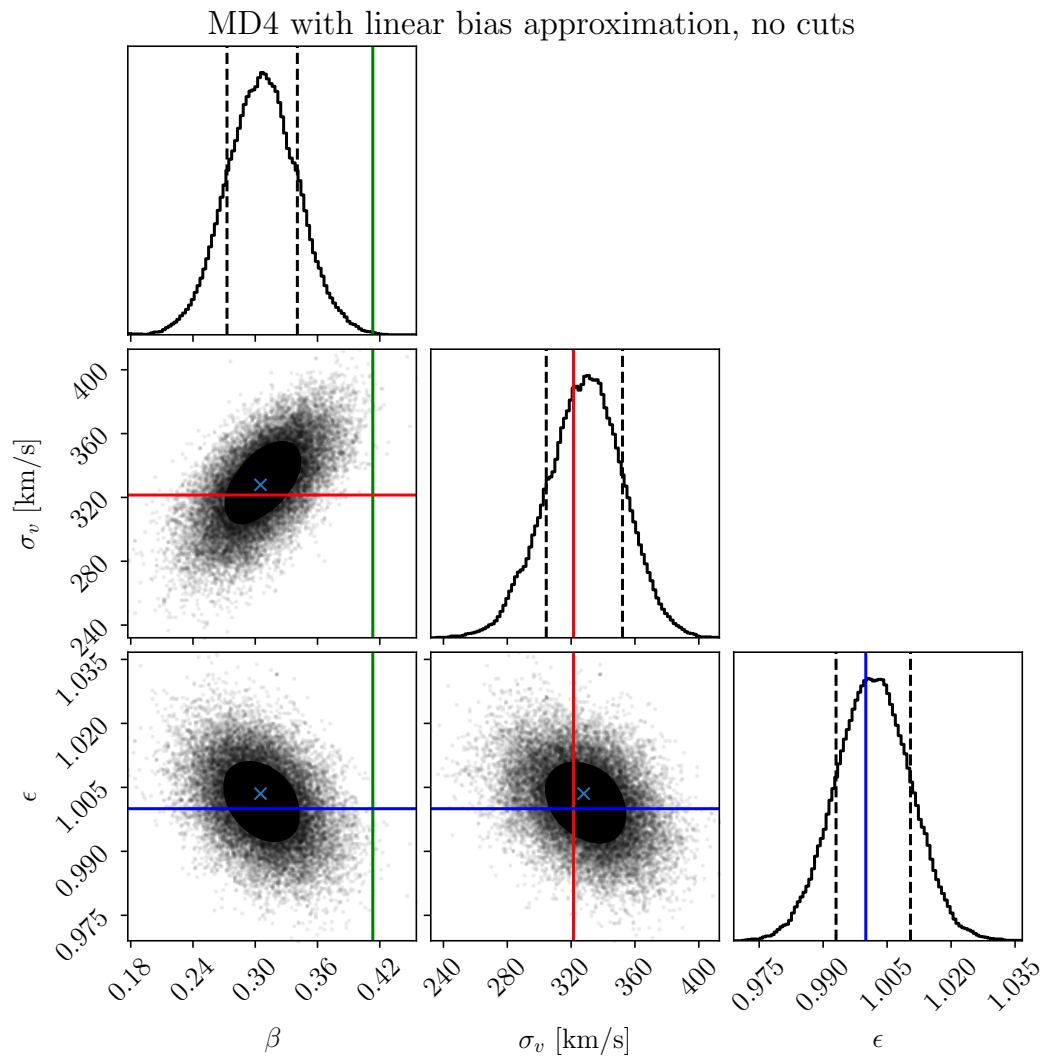


Figure 6.35: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory given by equation 5.12. The fiducial value for each parameter is overplotted.

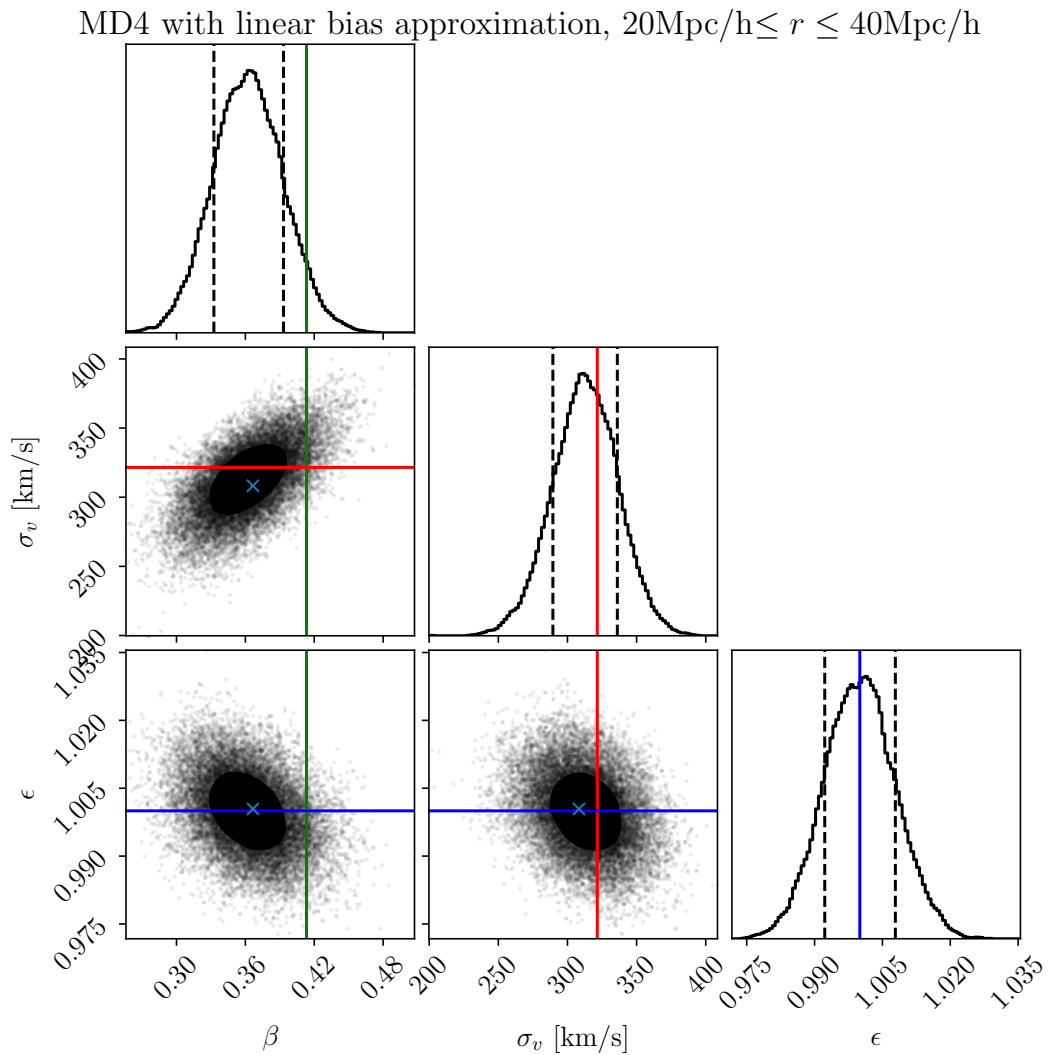


Figure 6.36: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. In this figure, only voids with effective radius $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ are considered. The fiducial value for each parameter is overplotted.

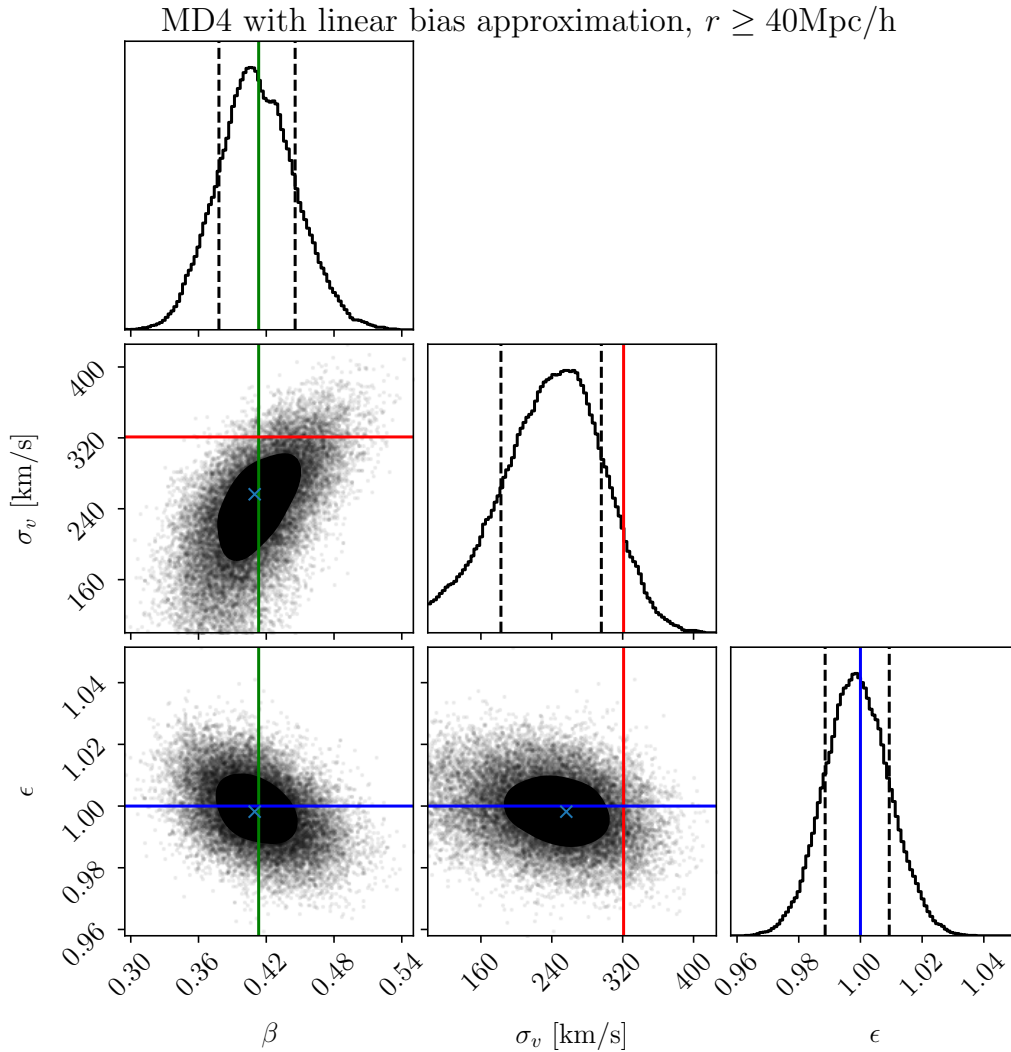


Figure 6.37: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory, given by equation 5.12. In this figure, only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered. The fiducial value for each parameter is overplotted.

6.3.2 Linear bias approximation with correction term

In order to study whether improvements can be made to the regular model derived from linear theory, a correctional term proposed by [4] was implemented, as described in section 5.1.3. The best fit values for the parameters ϵ , β and σ_v for all datasets tested using this model is shown in table 6.2. For the MD2 dataset, shown in figure 6.38, one

can see that this model does not fit the ϵ parameter as good as the regular model shown in figure 6.31. Although for this particular dataset, the model with the correction term performed slightly better when fitting β , it performed significantly worse when fitting ϵ . As ϵ is a parameter that is directly linked to the cosmology, this is the most important parameter in the model. Although β and ϵ are still within the 68% confidence interval, the correction term did not provide better fits than the regular model.

The same model, with the correctional term, is applied to the MD3 dataset without cuts, as shown in figure 6.39. Although it is slightly better than the regular model when fitting β for this dataset, as with the MD2 dataset, the model has a worse performance when fitting ϵ . This particular dataset contains a large number of small voids and may therefore be subject to non-linear effects. Therefore, as for the regular model, it may not be expected to perform as good on this particular dataset without cuts. However, the regular model still managed to capture the fiducial value for ϵ when applying the model to this dataset without cuts. This model, however, did not perform as well as the regular model in terms of sampling the fiducial value for ϵ . Figure 6.40 shows the model with the velocity correction for the MD3 dataset with the $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ cuts. This figure shows that for this dataset with cuts, the model was able to give an accurate sampling of β . As can be seen from figure 6.26, this model satisfyingly captured the amplitude of the numerically calculated velocity. The model with the velocity correction term, however, is not as sufficient when sampling the best parameters for ϵ . For the MD3 dataset with the $r \geq 40\text{Mpc}/h$ cut, as shown in figure 6.41, one can see that there is no significant improvement to the parameter fits in contrast to the regular model, where applying mass cuts seemed to have an effect where it reduced non-linear effects present. For this dataset with the given cut, the model does not perform as well as the regular model when sampling ϵ . Using this cut the sampling of β was slightly worse than for the $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ cuts applied. For the $r > 40\text{Mpc}/h$ cut, figure 6.26, illustrating the comparison between the two models with numerical data suggests that the regular model predicts a better velocity profile.

The model with the velocity correction term was also applied to the MD4 dataset without cuts. This is shown in figure 6.42. Here one can see that, as with the regular model for this dataset, β is not sampled accurately when there are no cuts. This is the same behavior as with the regular model for this dataset without cuts and may be attributed to the presence of a large number of small voids dominated by non-linear effects. The ϵ factor is not sampled as good as for the regular model. The MD4 dataset with the $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ cut applied can be seen in figure 6.43. Although the sampling of the β parameter is slightly better it is still outside the 68% confidence interval. As seen in figure 6.27, in contrast to the MD3 dataset with this cut applied, the model with the correction term did not capture the amplitude of the numerically calculated velocity significantly better than the regular model. The sampling of the ϵ parameter is also slightly worse than for the dataset without cuts. Figure 6.37 shows the model with the velocity correction term for the MD4 dataset only including voids with effective radius $r \geq 40\text{Mpc}/h$. The sampling of ϵ is approximately the same as for the other two figures for the MD4 dataset. Although, with this cut applied, the sampling of β is greatly improved. This behavior has proven to be consistent for both the MD3 and

Dataset	ϵ	β	σ_v [km/s]
MD2	0.9907 ± 0.0095	0.3090 ± 0.0320	392.109 ± 32.664
MD3 no cuts.	0.9893 ± 0.0094	0.3277 ± 0.0334	359.705 ± 27.378
MD3 $20\text{Mpc/h} \leq r \leq 40\text{ Mpc/h}$	0.9861 ± 0.0076	0.3492 ± 0.0294	359.092 ± 22.107
MD3 $r \geq 40\text{Mpc/h}$	0.9902 ± 0.0102	0.3920 ± 0.0352	299.747 ± 48.424
MD4	0.9913 ± 0.0089	0.3187 ± 0.0346	318.310 ± 23.532
MD4 $20\text{Mpc/h} \leq r \leq 40\text{ Mpc/h}$	0.9884 ± 0.0082	0.3787 ± 0.0314	305.919 ± 23.242
MD4 $r \geq 40\text{ Mpc/h}$	0.9851 ± 0.0105	0.4361 ± 0.0348	240.308 ± 55.041

Table 6.2: Best fit values for the parameters ϵ , β and σ_v for the different datasets using a linear bias approximation and modelling the velocity using an approximation term, given in equation 5.13.

MD4 datasets, using both the models, with and without the velocity correction term. The sampling of ϵ remains almost the same with or without cuts, however, the regular model sampled it closer to the fiducial value for all the datasets. For the model with the correction term, the sampling of ϵ is biased towards sampling a lower value than the fiducial value. The β parameter on the other hand is highly affected by the cuts.

MD2 with linear bias approximation
and Achitouv (2017) term, no cuts

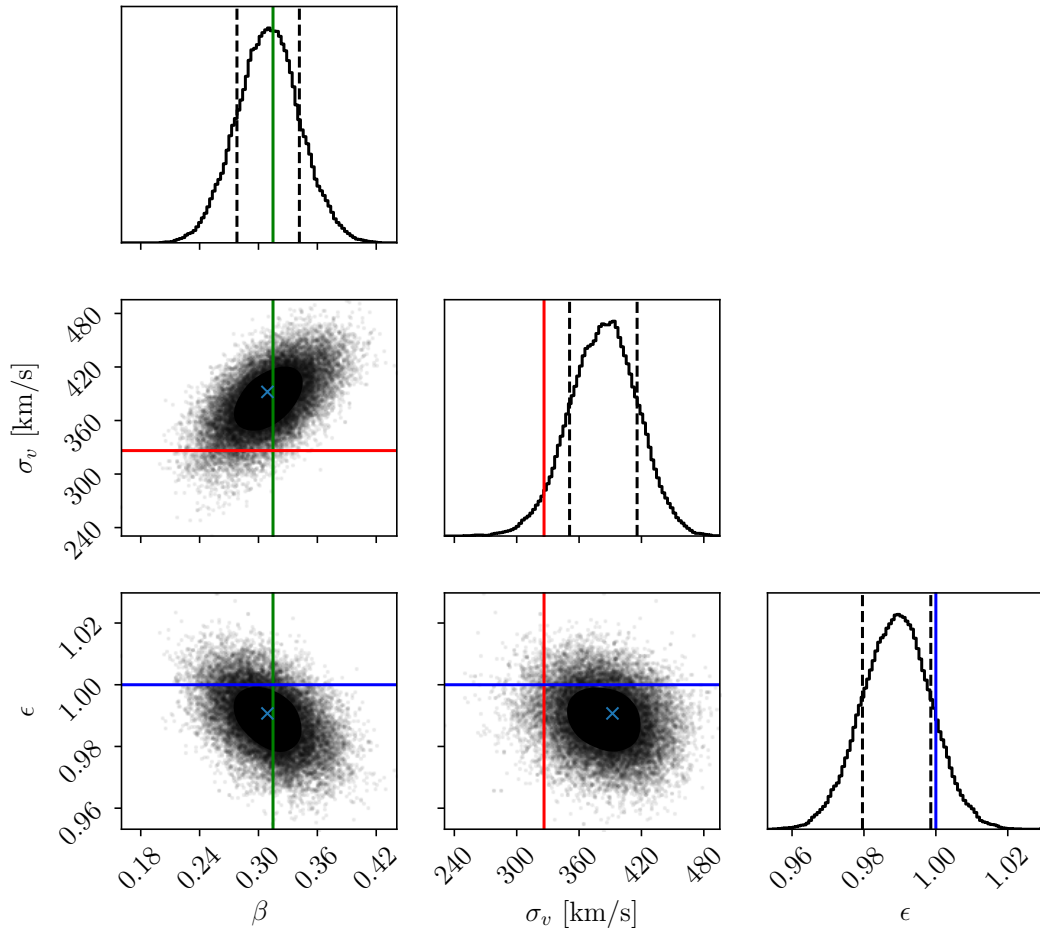


Figure 6.38: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 2 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13.

MD3 with linear bias approximation
and Aчитouv (2017) term, no cuts

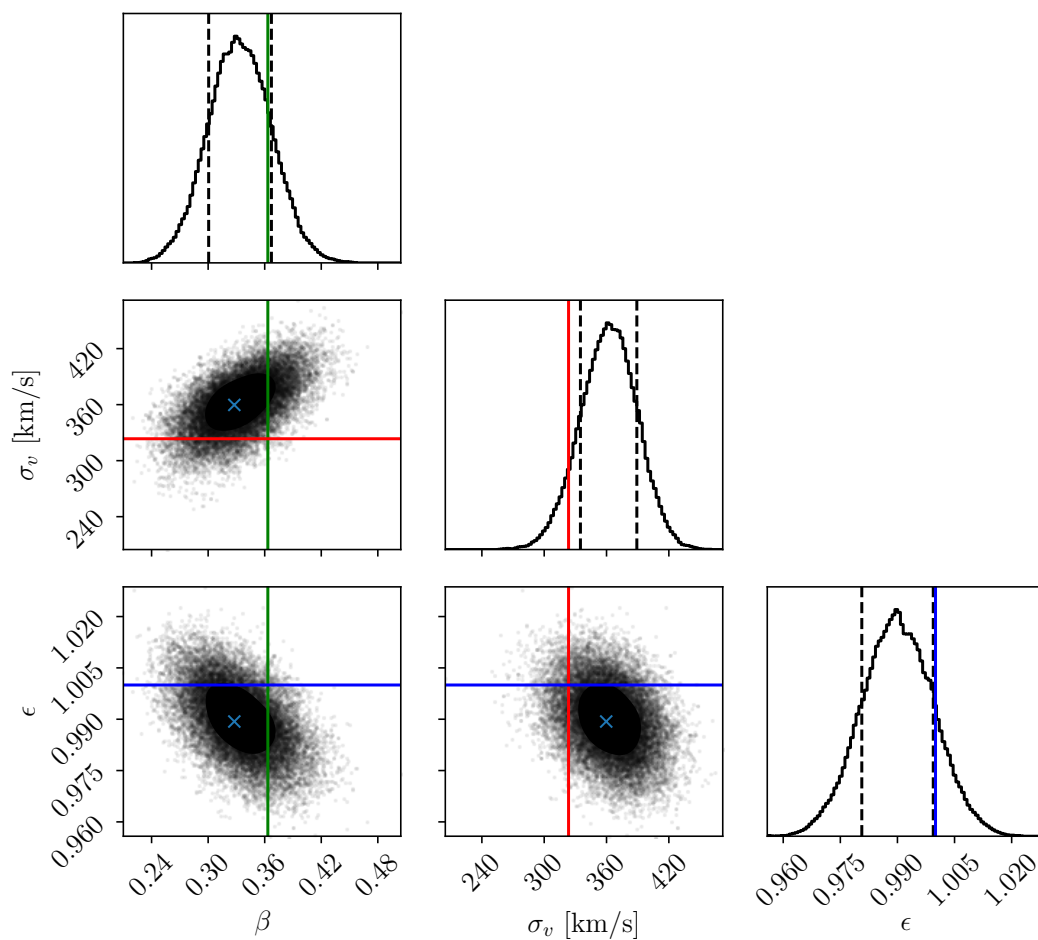


Figure 6.39: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13.

MD3 with linear bias approximation
and Aчитouv (2017) term, $20\text{Mpc}/h < r < 40\text{Mpc}/h$

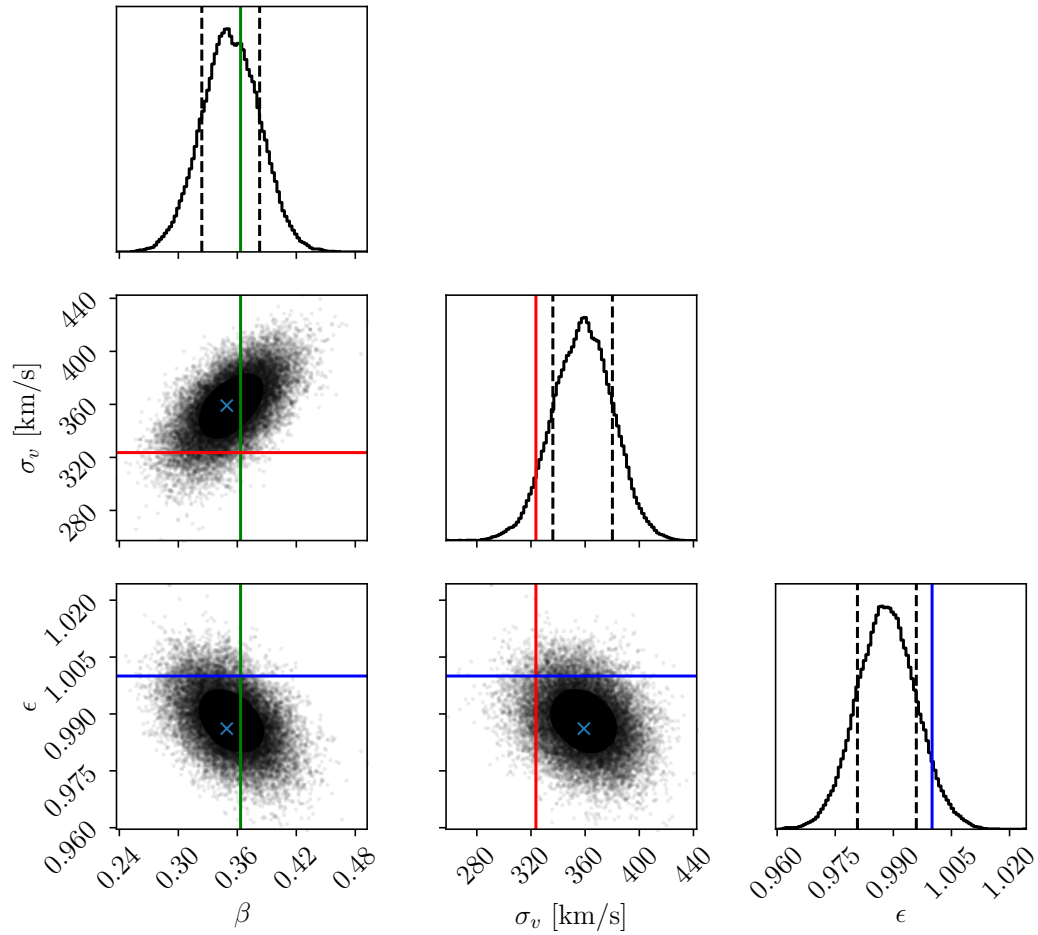


Figure 6.40: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13. In this figure, only voids with effective radius $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ are considered.

MD3 with linear bias approximation
and Aчитouv (2017) term, $r > 40\text{Mpc}/h$

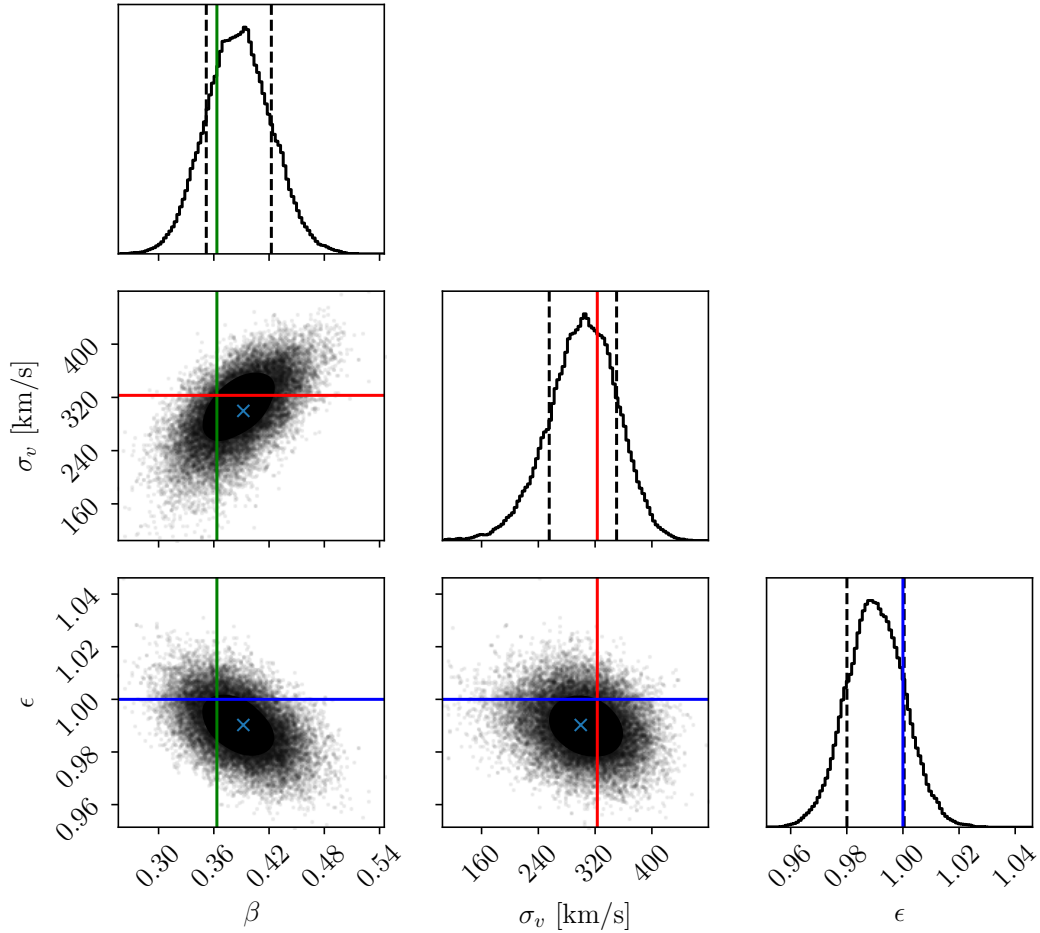


Figure 6.41: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 3 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13. In this figure only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered.

MD4 with linear bias approximation
and Aчитouv (2017) term, no cuts

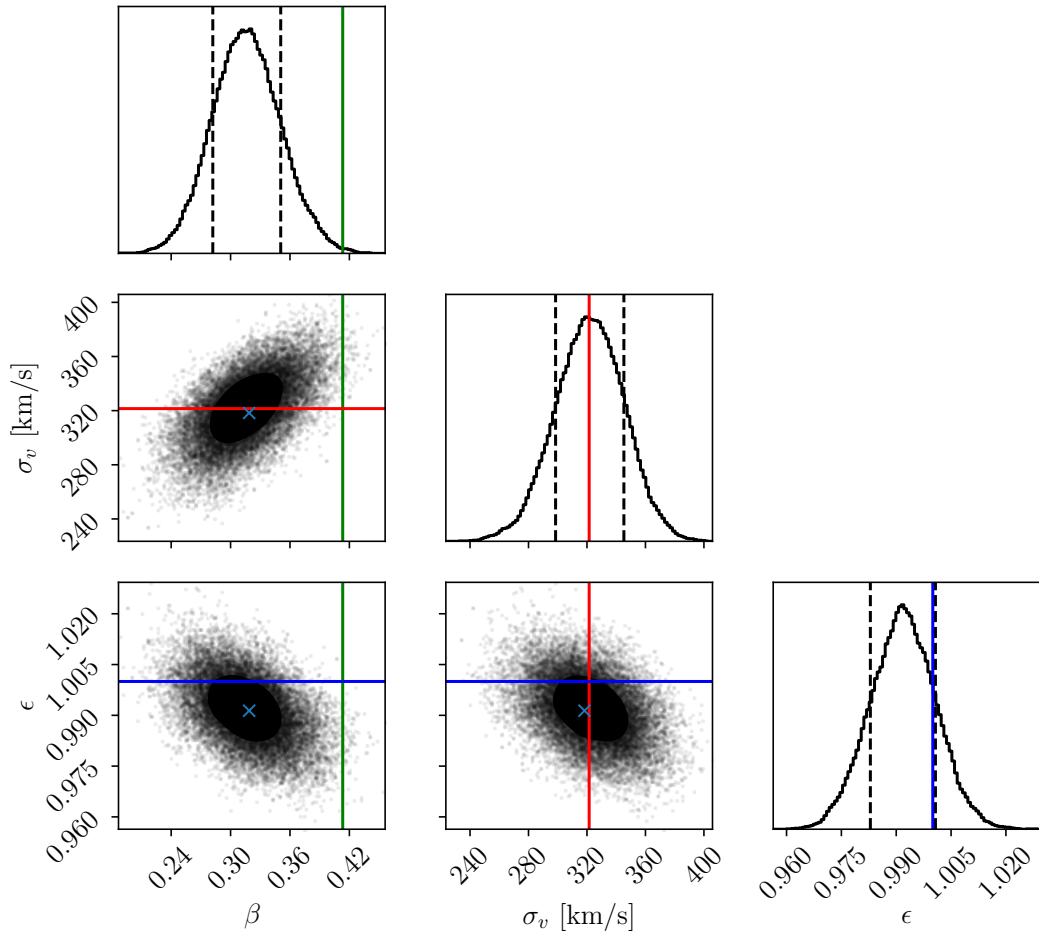


Figure 6.42: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13.

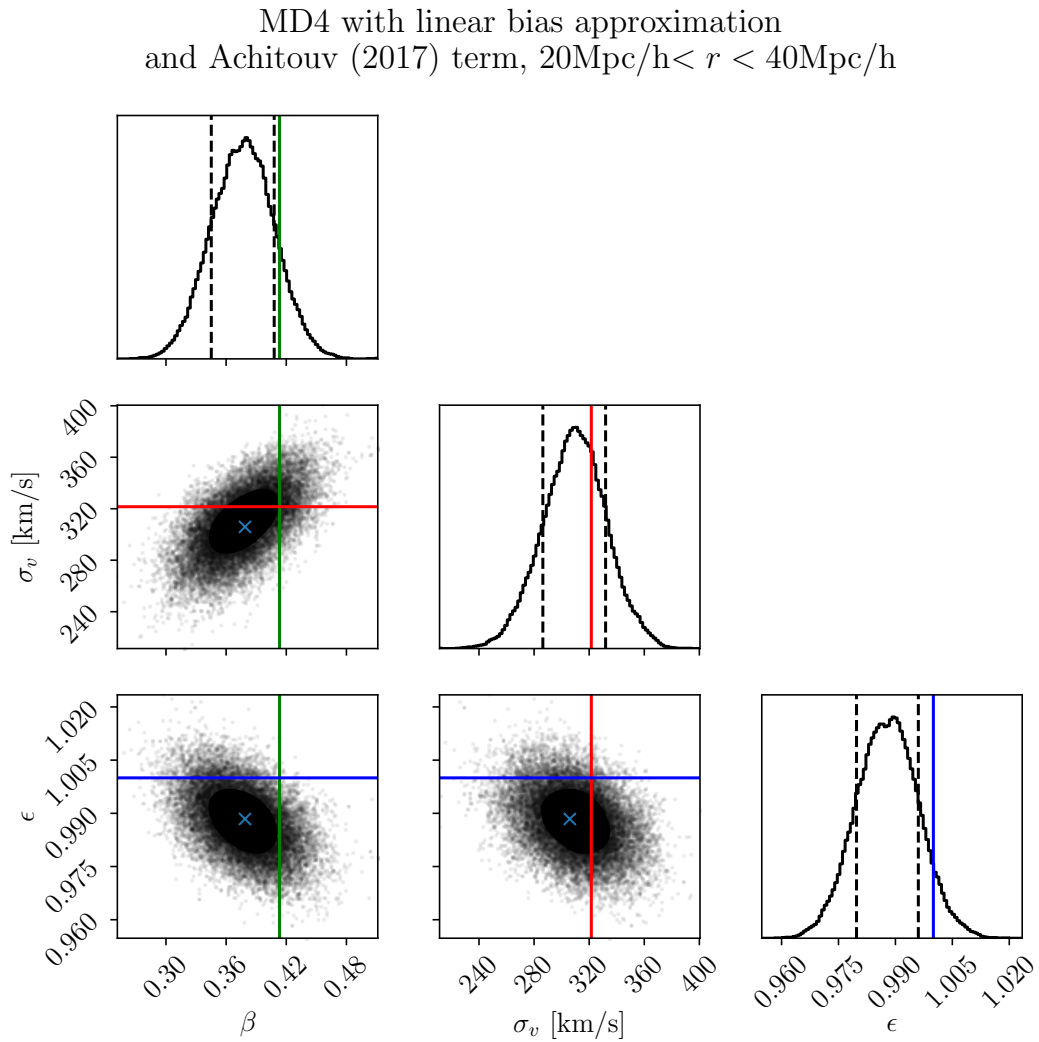


Figure 6.43: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13. In this figure only voids with effective radius $20\text{Mpc}/h \leq r \leq 40\text{Mpc}/h$ are considered.

MD4 with linear bias approximation
and Aчитouv (2017) term, $r > 40\text{Mpc}/h$

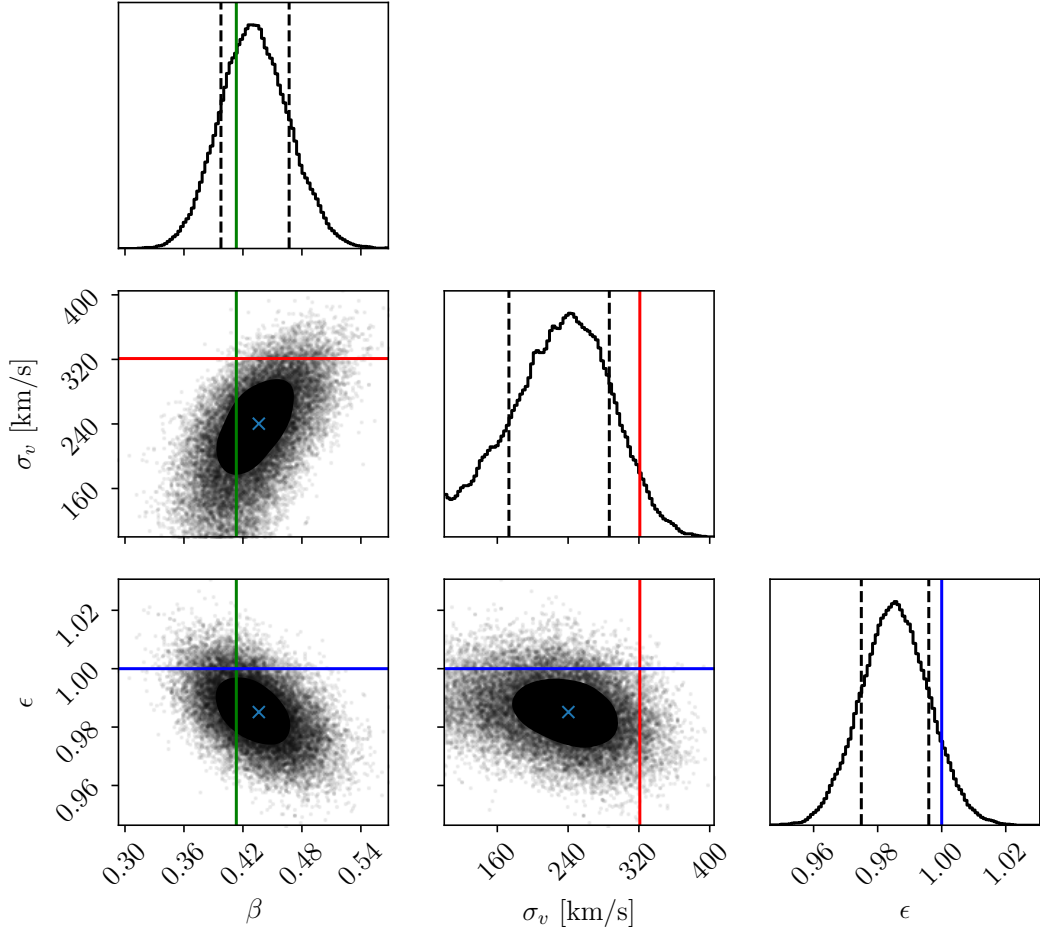


Figure 6.44: Figure showing showing parameters fits to β , σ_v and ϵ for the Multidark 4 dataset using a linear bias approximation and modelling the velocity using linear theory and a velocity correction proposed by [4], given by equation 5.13. In this figure only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered.

6.3.3 Dark matter profile fits

Using a dark matter density profile, described in section 5.6, parameter fits for $f\sigma_8$, σ_v , r_{scale} and ϵ was calculated. For these fits, the $20\text{Mpc}/h < r < 40\text{Mpc}/h$ cuts were excluded. For the MD2 dataset, this is illustrated in figure 6.45. From this figure, one can see that the model accurately samples the ϵ factor. However, $f\sigma_8$, which is the

other important parameter in this model, is not accurately sampled. From figure 6.19, which shows the dark matter profile used for the parameter fits for the MD2 dataset, one can see that the peaks of the δ_{dm} profile coincides almost at the same radius as the peak of δ_g . From the fits to r_{scale} , one can see that this is sampled to have a maximum likelihood value of approximately 1.0. This qualitative measurement suggests that the model accurately picks the correct value for scaling the radius of the dark matter density profile for this dataset. The sampling of $f\sigma_8$ suggests that the model is not sufficient when sampling the correct amplitude of the dark matter density profile. This may be expected since the dark matter density profile for the given dataset was not used.

The parameter fits for the MD3 dataset, using a dark matter density profile and no cuts, is shown in figure 6.46. For this dataset, the model accurately samples ϵ . As for the MD2 dataset, $f\sigma_8$ is not sampled accurately. As can be seen from the dark matter density profile used in this dataset, shown in figure 6.20, the peak of δ_g is shifted towards slightly smaller radii than that of δ_{dm} . From the fits to r_{scale} , one can see that the model places the most likely value for this parameter at slightly less than 1.0. This will shift δ_{dm} towards smaller radii, in which the qualitative measurement from figure 6.20 suggests. However, as with the MD2 dataset, the model does not provide sufficient fits to the amplitude as measured by the $f\sigma_8$ parameter. In order to test if this model was subject to errors introduced by non-linear behavior, as for the linear bias approximation, the $r > 40\text{Mpc}/h$ cut was applied. Parameter fits for this dataset and cut, using the model with a dark matter overdensity profile, is shown in figure 6.47. As for the same dataset without cuts, the model accurately samples the ϵ parameter. $f\sigma_8$ on the other hand is sampled closer to the fiducial value than for the MD3 dataset without cuts. But, it is still not sampled accurately enough to provide reliable fits. One can imagine applying even more conservative cuts could improve the results. But this cut will remove even more voids from the catalogue, making the number of void-halo pairs decrease, which in turn will result in more noise when numerically calculating the two-point correlation function. As seen from δ_g for the MD3 dataset, illustrated in figure 6.17, applying cuts will shift the peak of the overdensity profile towards higher radii. The fits to the r_{scale} parameter shows that the best fit for this parameter is slightly larger than 1.0. This suggests that the model captures the scaling of δ_{dm} in the radial direction but it is not sufficient when scaling the amplitude with $f\sigma_8$.

Using the MD4 dataset, parameter fits using the velocity predicted by linear theory with a dark matter overdensity profile were calculated. For this dataset without cuts, the parameter fits are illustrated in figure 6.48. As for the other datasets tested using a dark matter overdensity profile, the parameter fits to ϵ are satisfying. The $f\sigma_8$ parameter on the other hand, as with the MD3 dataset without cuts, is sampled at a significantly lower value than the fiducial value. Figure 6.21 suggests that δ_{dm} should be shifted towards lower radii. The r_{scale} parameter reflects this in that the most likely value is sampled at a value lower than 1.0. The same model was tested on the MD4 dataset with the $r > 40\text{Mpc}/h$ cuts applied. This is shown in figure 6.49. As with all the other tests applied using this model, the sampling of the ϵ parameter turned out to be satisfying. Compared to the MD4 dataset without cuts, the sampling of $f\sigma_8$ is better but it is still not satisfying. This behavior was also found for the MD3 dataset with and without

cuts. This may be attributed to non-linear effects present in smaller voids. However, as previously stated, applying more conservative cuts than the $r < 40\text{Mpc}/h$ cut will drastically remove a significant amount of void-halo pairs, making the cross-correlation function itself subject to more noise. The model did however manage to sample both the β and ϵ values correctly when a linear bias assumption was applied. This is in contrast to using the dark matter overdensity profile when applying the $r > 40\text{Mpc}/h$ cuts. This would in turn suggest that these cuts should be sufficient. It may instead suggest that attempting to tune a dark matter density profile derived from another dataset can prove problematic for accurate sampling the $f\sigma_8$ parameter. The r_{scale} parameter is sampled with a best fit value larger than 1.0. This is reasonable as δ_g for the MD4 dataset shows that the overdensity profile gets shifted towards larger radii with this cut. This model was able to fit the most important parameter linked to the cosmology, namely ϵ . However using a dark matter overdensity profile derived from another dataset, these results may suggest that attempting to adjust the profile itself is not sufficient for accurate sampling of the $f\sigma_8$ parameter.

MD2 with dark matter profile.

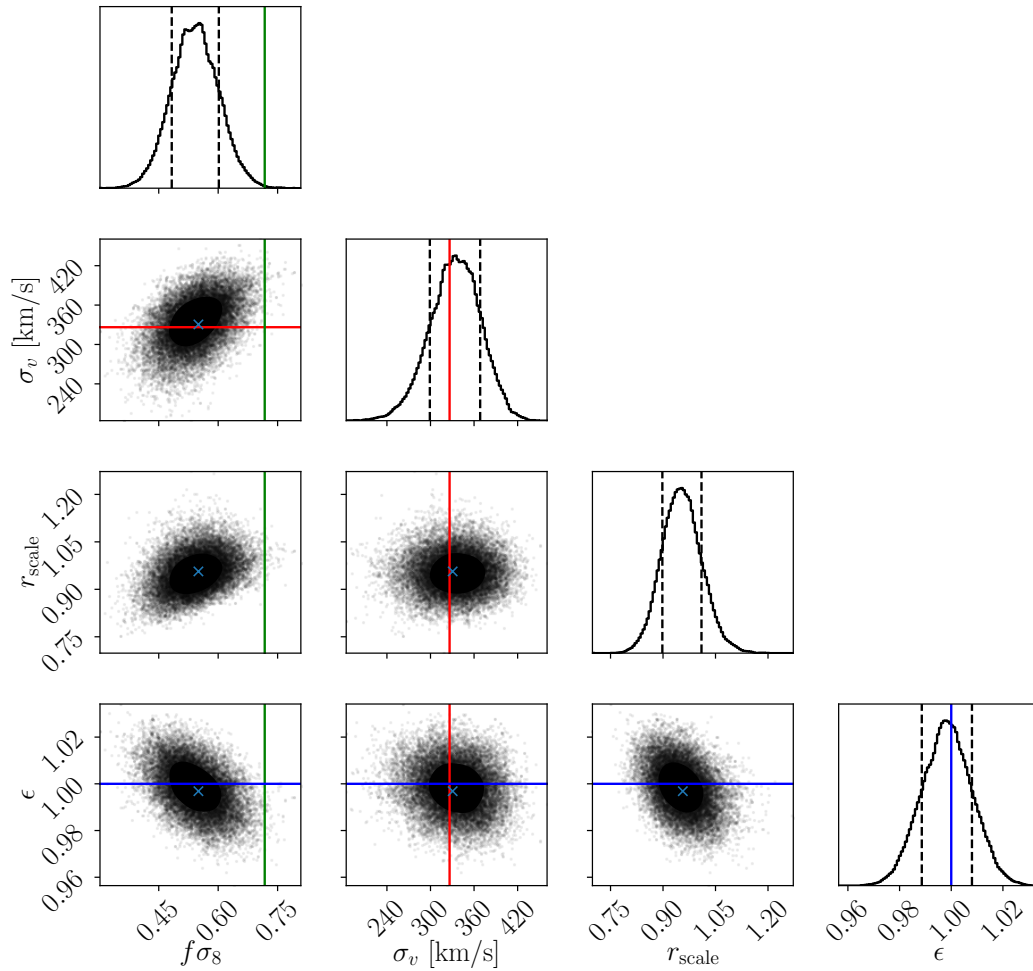


Figure 6.45: Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 2 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given in equation 5.12

MD3 with dark matter profile.

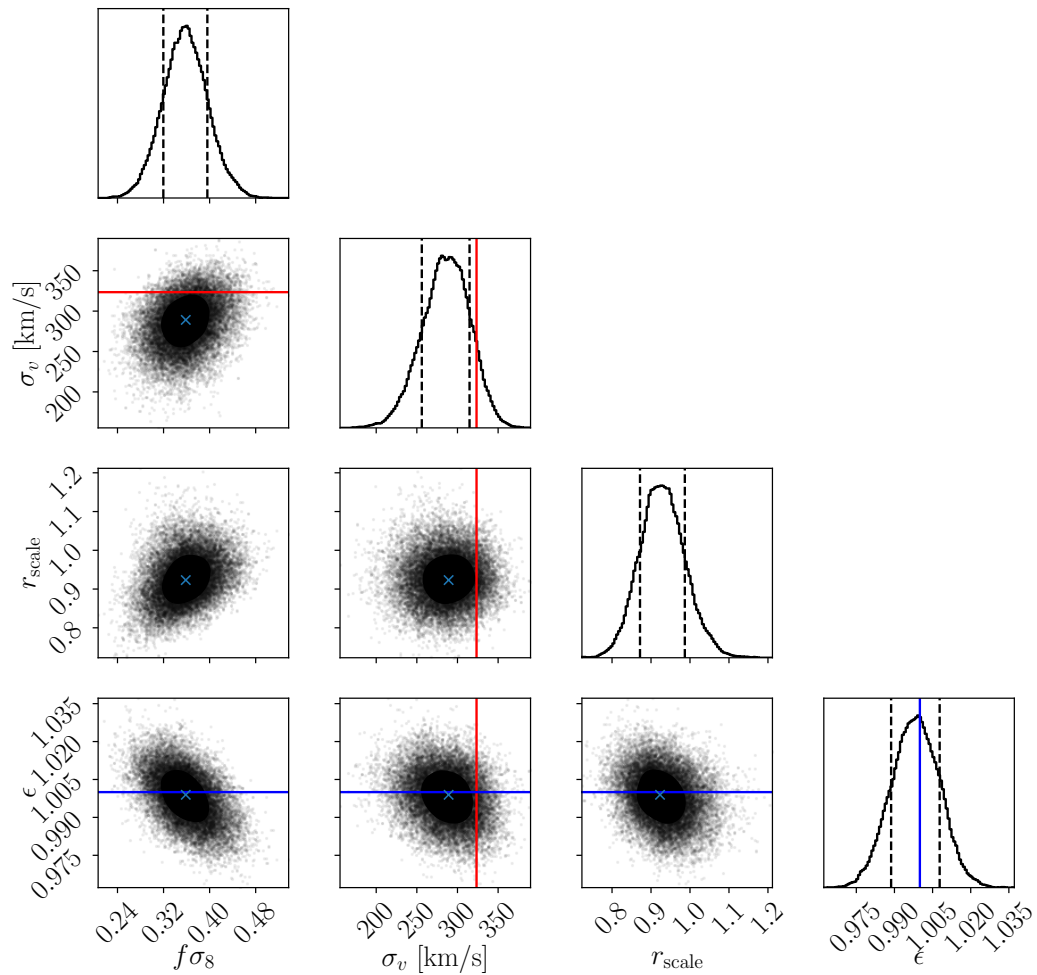


Figure 6.46: Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 3 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given by equation 5.12

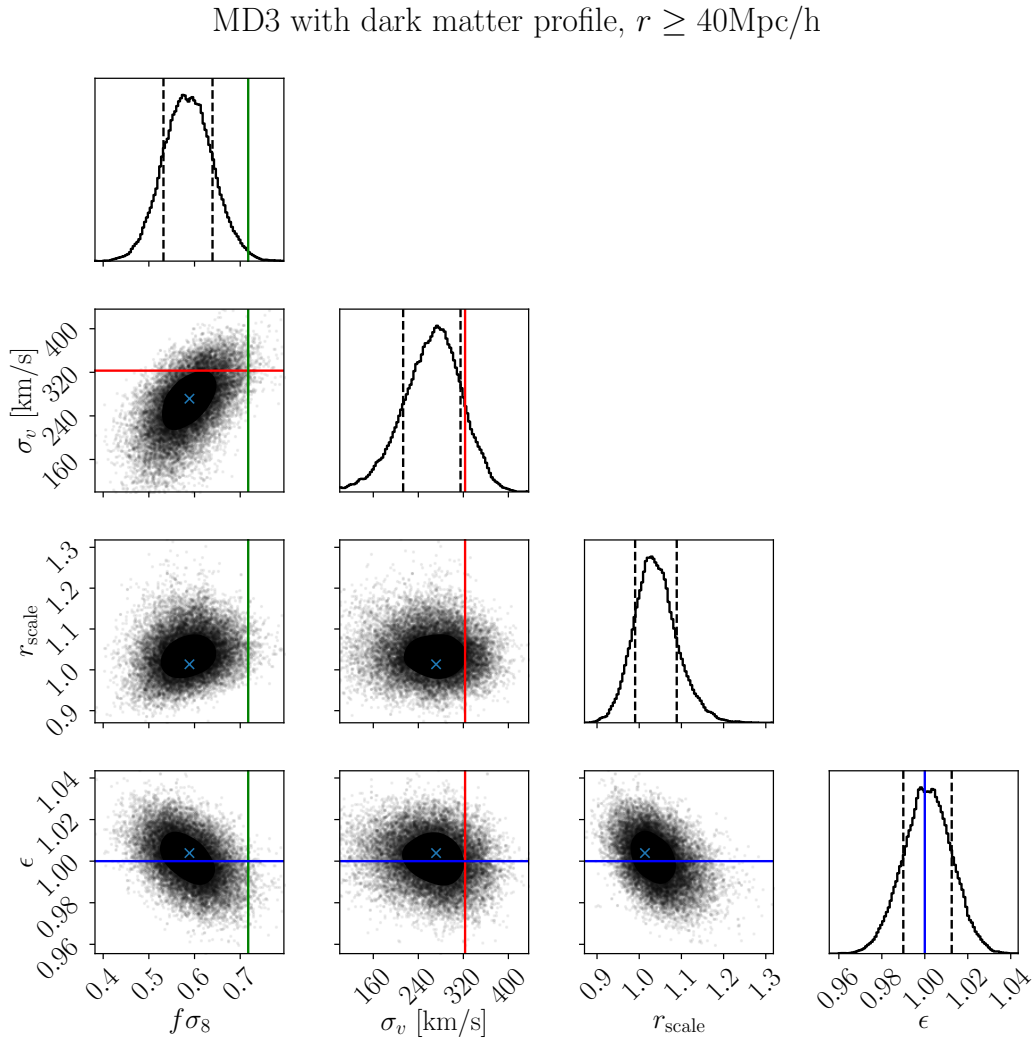


Figure 6.47: Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 3 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given in equation 5.12. In this figure only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered.

MD4 with dark matter profile.

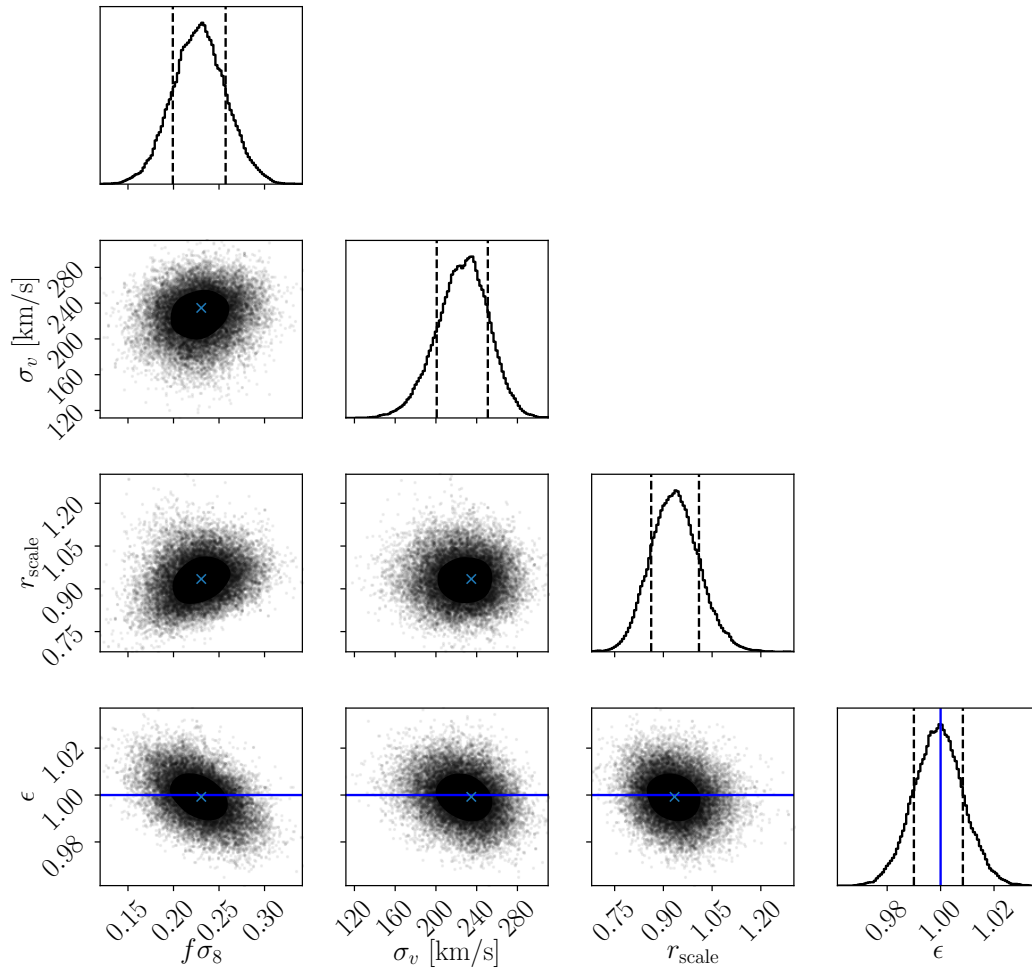


Figure 6.48: Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 4 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given in equation 5.12

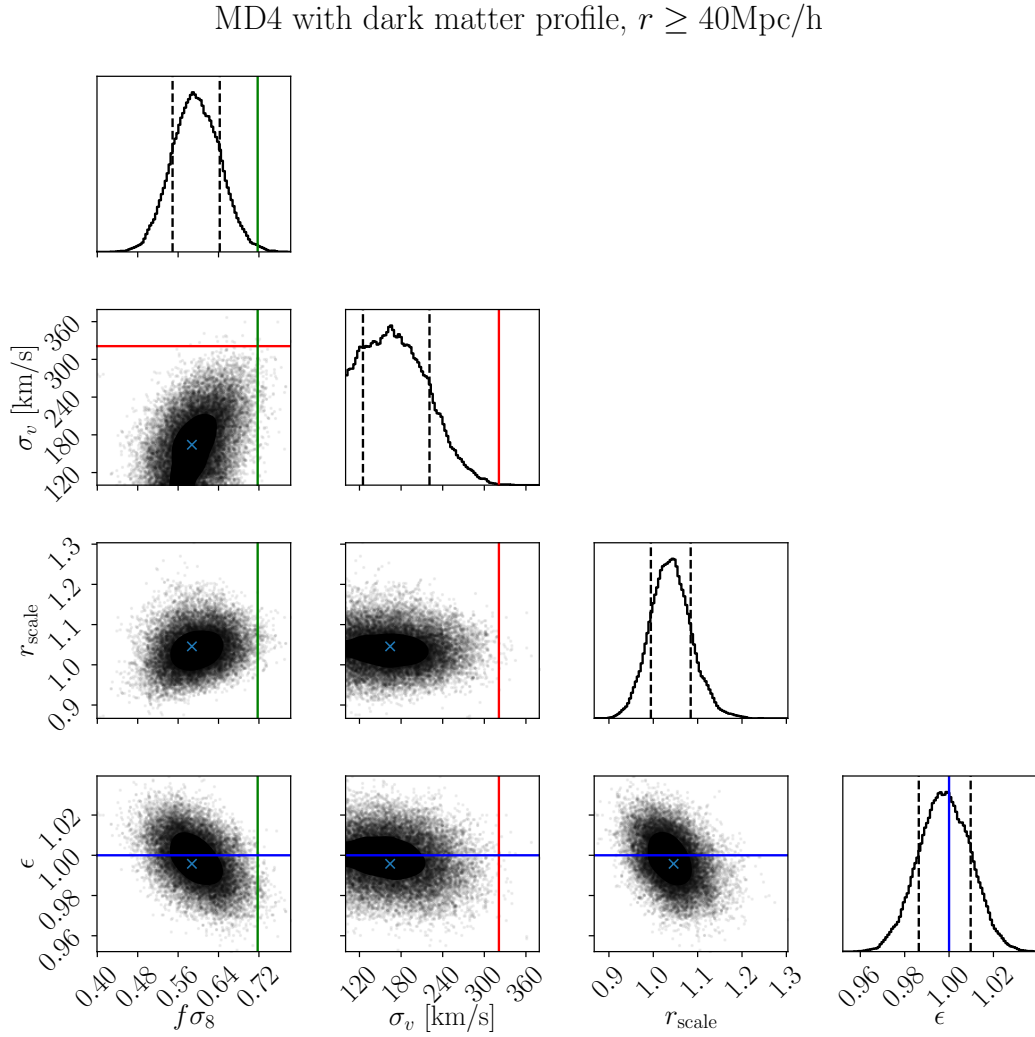


Figure 6.49: Figure showing showing parameters fits to $f\sigma_8$, σ_v , r_{scale} and ϵ for the Multidark 4 dataset using a dark matter profile, as described in section 5.6, and modelling the velocity using linear theory given in equation 5.12. In this figure only voids with effective radius $r \geq 40\text{Mpc}/h$ are considered.

6.3.4 Correlation function void-galaxy ξ_{vg}

Using the best fit parameters for the linear bias approximation, with and without the additional correction term, given in tables 6.1 and 6.2, a comparison for the calculated quadrupole moment $\xi_2^s(s)$ of the correlation function $\xi_{\text{vg}}^s(s, \mu)$ was calculated. The correlation function calculated is the streaming model given in equation 5.39. $\xi^s(s, \mu)$

Dataset	ϵ	$f\sigma_8$	σ_v [km/s]	r_{scale}
MD2 no cuts.	0.9968 ± 0.0096	0.5501 ± 0.0597	330.535 ± 35.084	0.95657 ± 0.05719
MD3 no cuts.	0.9989 ± 0.0094	0.3585 ± 0.0390	289.107 ± 29.540	0.92314 ± 0.05873
MD3 $r \geq 40\text{Mpc/h}$	1.0003 ± 0.0112	0.5888 ± 0.0541	271.288 ± 51.705	1.01360 ± 0.05174
MD4	0.9992 ± 0.0092	0.2304 ± 0.0292	234.701 ± 25.439	0.93445 ± 0.07486
MD4 $r \geq 40$ Mpc/h	0.9957 ± 0.0116	0.5872 ± 0.0469	164.232 ± 44.885	1.04583 ± 0.04659

Table 6.3: Best fit values for the parameters ϵ , σ_v , $f\sigma_8$ and r_{scale} for the different datasets using a dark matter density profile.

itself is calculated using full expression given by equation 5.29. This equation was implemented with and without the additional correction term proposed by [4] using the best-fit parameters for the respective model and dataset. For all figures, the green line is the quadrupole numerically calculated by the pyCUTE code. The blue and orange lines are the quadrupole calculated using the regular velocity predicted by linear theory given in equation 5.12, and the velocity predicted using the correction term, given in equation 5.13, respectively. The quadrupole of the cross-correlation function for the MD2 dataset is shown in figure 6.50. Due to a lower sample size of void and halo pairs in this dataset, there is a lot of noise present in the numerically calculated quadrupole. This also makes it hard to qualitatively judge the performance of the two models by comparing the theory curves with the numerical curve. This may also suggest that not much emphasis should be put into the model fits for this dataset. From figures 6.31 and 6.38, one can see that although the regular model performed much better when fitting ϵ , the model with the correction term performed slightly better when fitting β . The differences in the calculated quadrupole of the correlation function between the two models are most evident between $20\text{Mpc/h} < s < 60\text{Mpc/h}$. The regular model qualitatively captures the central peak of the quadrupole in a more accurate fashion.

The quadrupole $\xi_2^s(s)$ for the MD3 dataset without cuts is shown in figure 6.51. For this dataset without cuts, the parameter fits suggested that the presence of small voids could introduce non-linear effects not accurately approximated by neither of the two velocity models. However, as the regular velocity model performed better at fitting the ϵ parameter, the model with the correction term proved slightly better at fitting β for this dataset without cuts. The β parameter acts as a normalization for the correlation function. The amplitude of the central peak is captured better by the model with the correction term. In contrast, the regular model captures the behavior of the numerically calculated quadrupole better for scales approximately $20\text{Mpc/h} < s < 30\text{Mpc/h}$. On scales $s > 60\text{Mpc/h}$ the two models look almost identical. Overall, the two calculated models look similar, but the model with the velocity correction term has a slightly better normalization on the central peak. The quadrupole $\xi_2^s(s)$ for the MD3 dataset, with the cuts only including voids with effective radii $r > 40\text{Mpc/h}$, is shown in figure

6.52. For this dataset, the regular model provided the better fits for both the ϵ and β parameters. The figure illustrating the $\xi_2^s(s)$ for this dataset with the given cut, shows that the regular model qualitatively gives the best fits to the correlation function. On small scales $s < 20\text{Mpc}/h$, the two models look almost identical, but for $s > 20\text{Mpc}/h$, the regular model qualitatively resembles the numerically calculated correlation function the closest. The amplitude of the quadrupole predicted by the model with the correction term has a lower amplitude than the numerically calculated on these scales. With these cuts however one can also see that the correlation function contains a lot more noise. This is due to applying the cuts will result in a reduced number of void-halo pairs. This issue could be addressed by applying different cuts gaining more pairs from the dataset, but still retaining voids where linear theory is still a sufficient approximation.

The quadrupole moment for the MD4 dataset without cuts is shown in figure 6.53. With this dataset, for both of the models, the sampling of β proved to not match the fiducial value for the dataset accurately. The regular model provided a better sample of the ϵ parameter. For small scales, one can see that the model with the correction term qualitatively resembles the numerically calculated quadrupole better. The regular model has an amplitude that is too large compared with the numerical quadrupole for both the first trough and the first peak. Figure 6.54 shows $\xi_2^s(s)$ for the MD4 dataset with the $r > 40\text{Mpc}/h$ cut applied. As with the calculated quadrupole for the MD3 dataset, the regular model qualitatively resembles the numerically calculated quadrupole better on scales $20\text{Mpc}/h < s < 60\text{Mpc}/h$. The model with the velocity correction term does not capture the amplitude of the numerically calculated model. The parameter fits for both models with these cuts showed that the regular model provided good samples for both ϵ and β . The model with the correction term provided good fits for β . It did however not perform as good when fitting ϵ . With the cuts applied, as for the MD3 dataset with these cuts, the number of void-halo pairs gets drastically reduced. This is evident in the numerically calculated quadrupole. It would therefore be useful for future testing to apply more conservative cuts in order to include more pairs while still staying inside the regime where linear theory is a sufficient approximation.

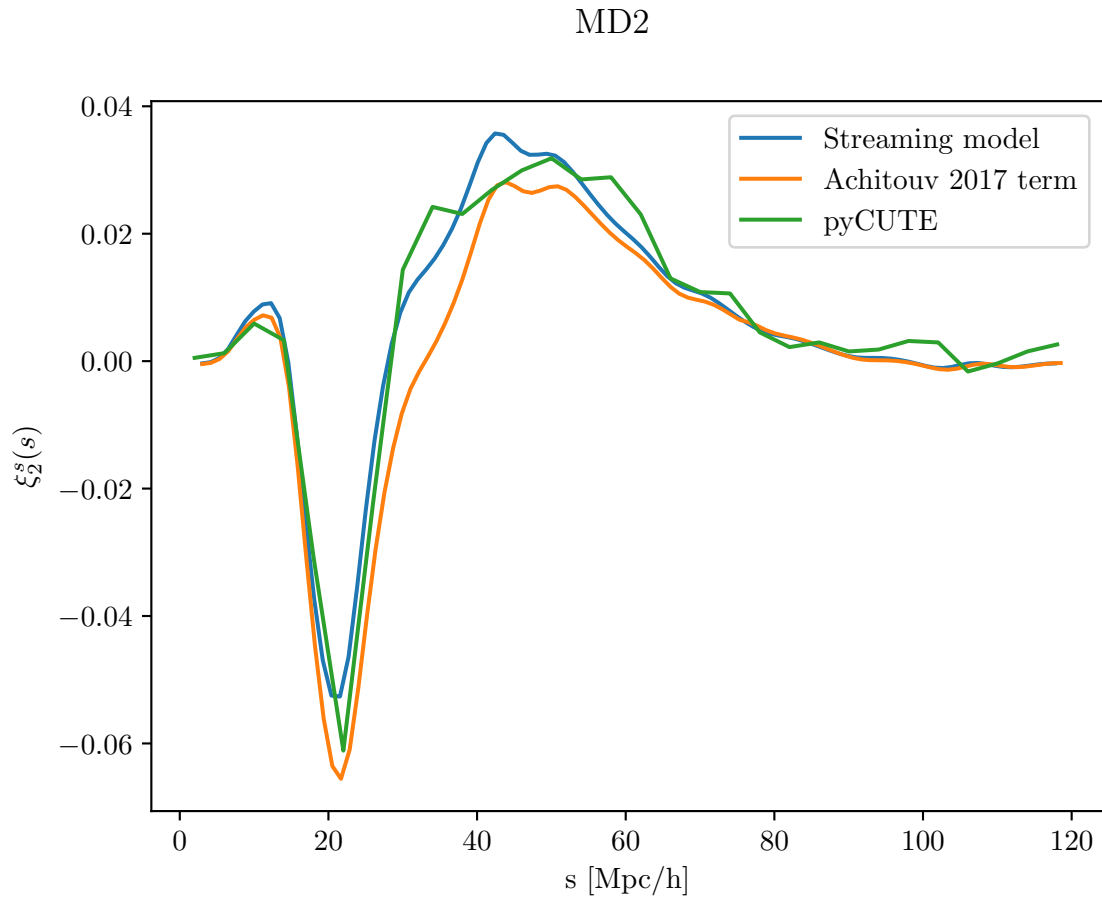


Figure 6.50: Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD2 dataset.

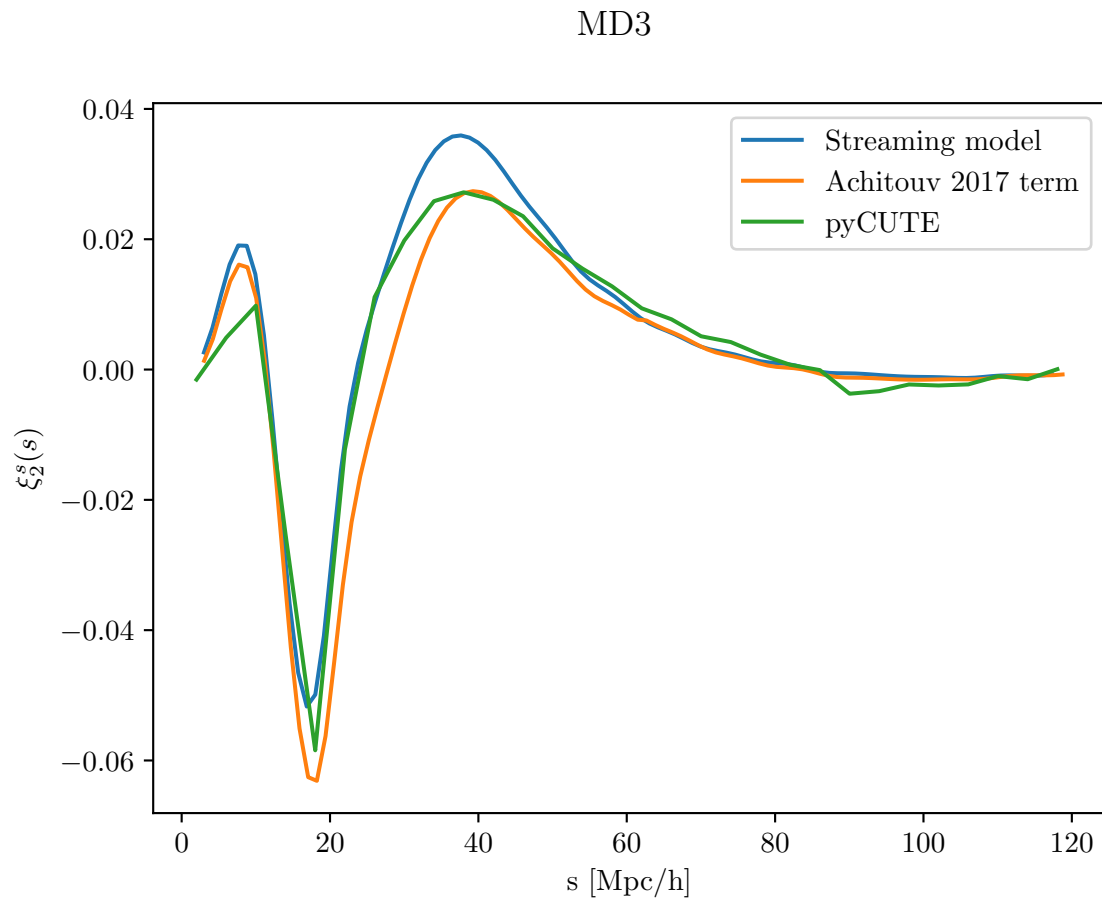


Figure 6.51: Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD3 dataset.

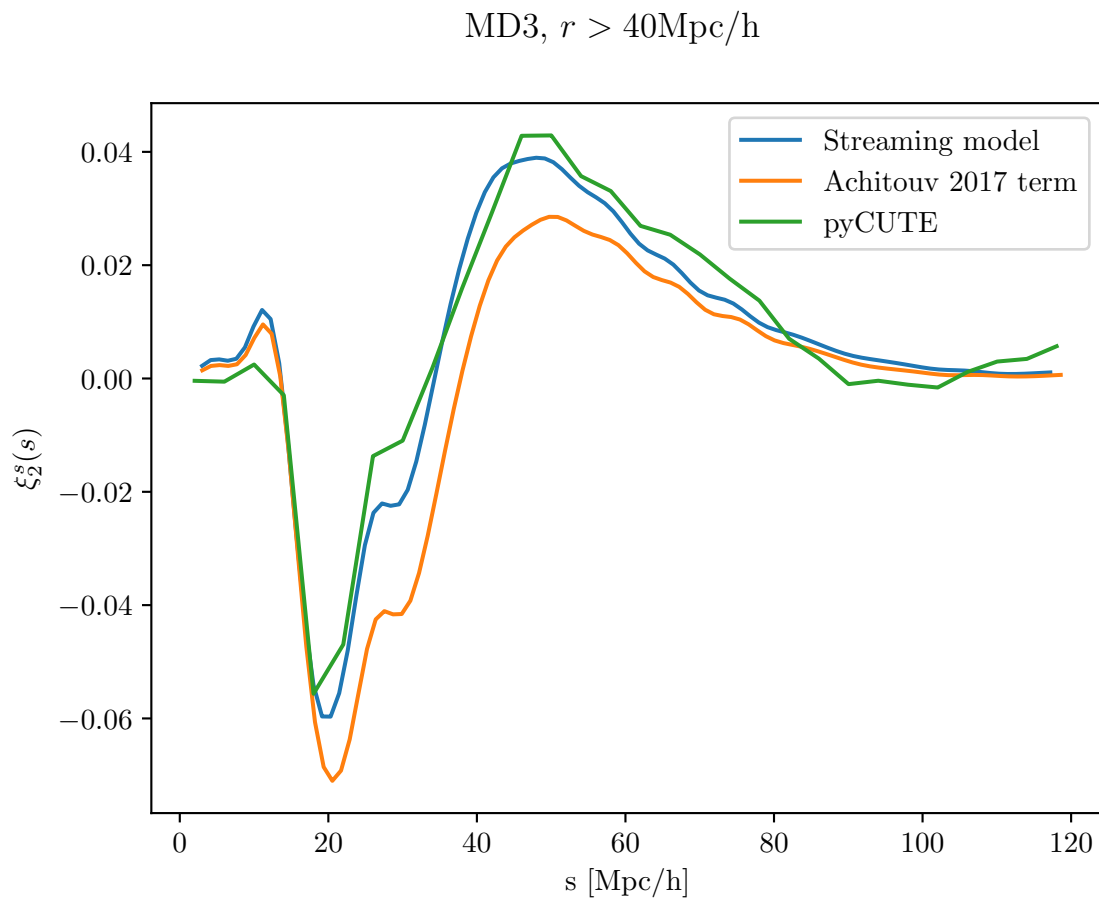


Figure 6.52: Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD3 dataset. Here only voids with effective radius $r \geq 40\text{Mpc/h}$ is included.

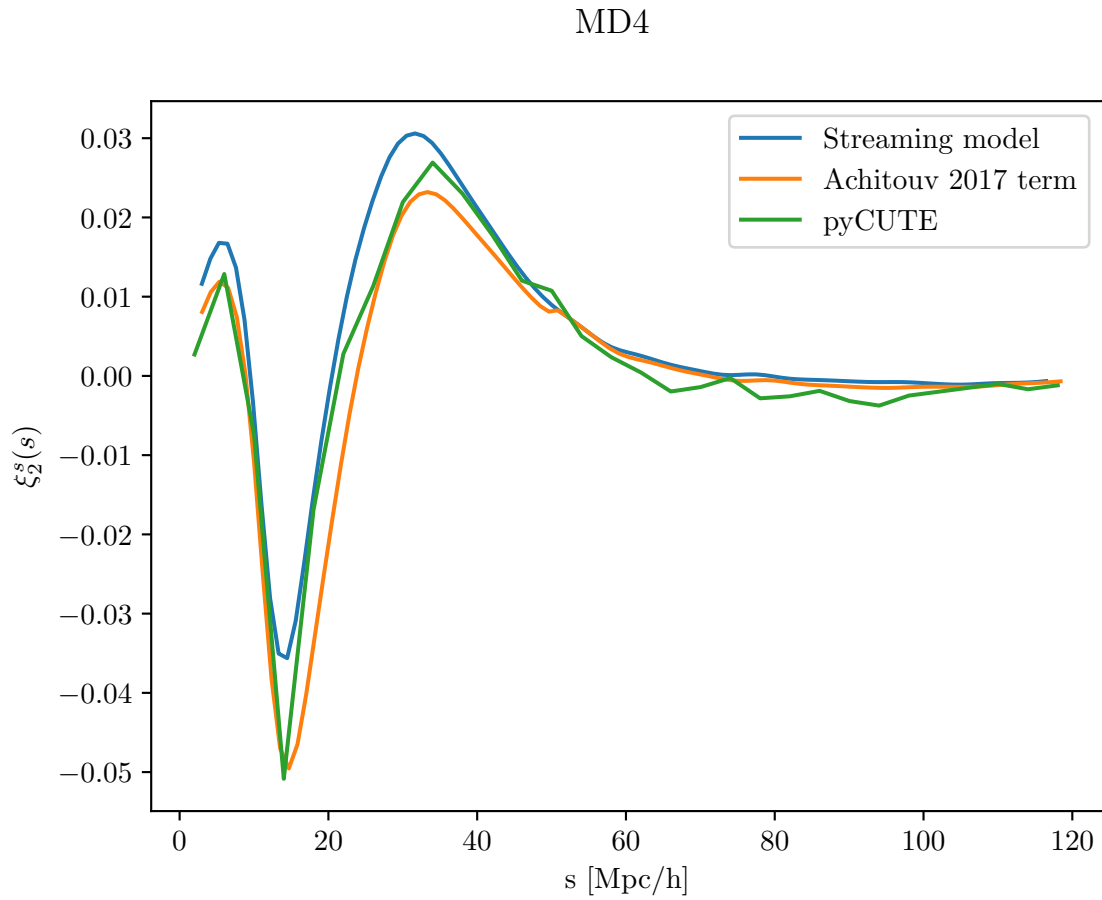


Figure 6.53: Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD4 dataset.

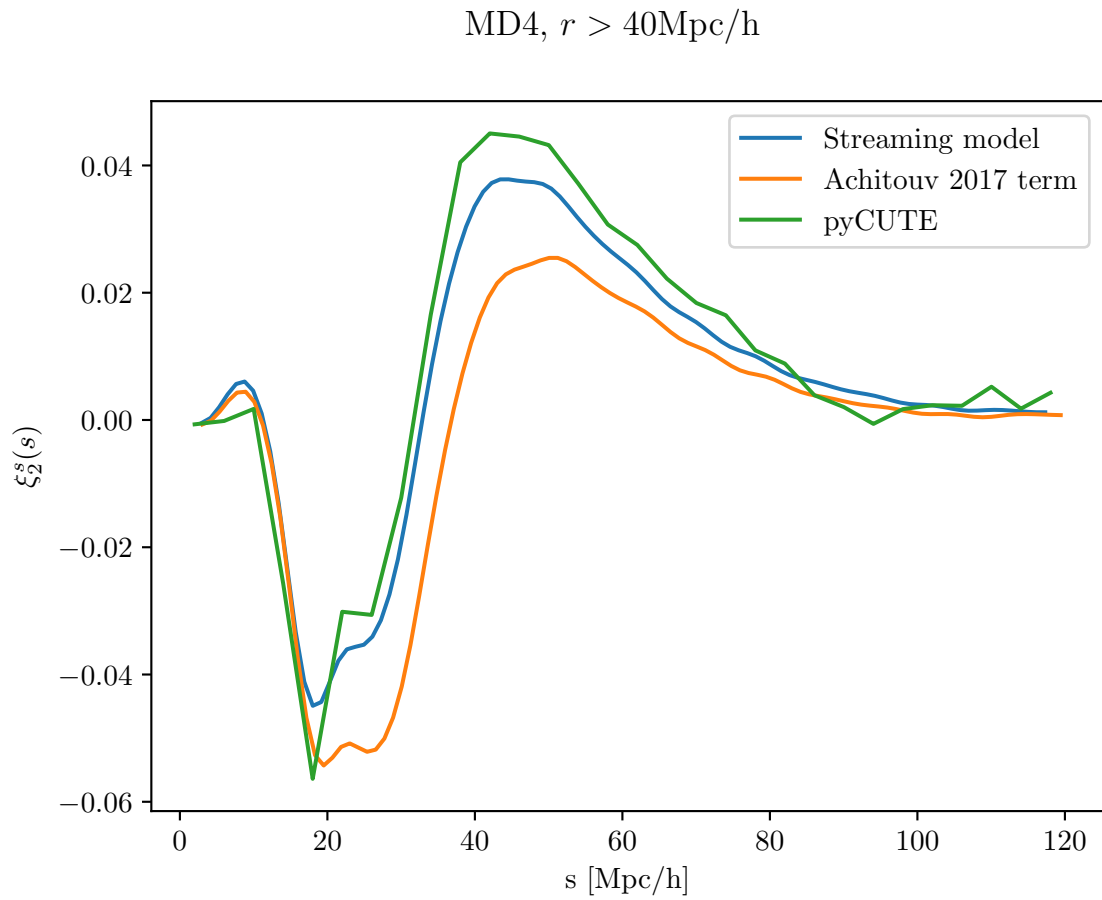


Figure 6.54: Figure showing the quadrupole of the cross correlation function $\xi_2^s(s)$ using the best fit parameters for the streaming model with both the regular velocity model and the model with the velocity correction term, given in tables 6.1 and 6.2 respectively, for the MD4 dataset. Here only voids with effective radius $r \geq 40\text{Mpc/h}$ is included.

Chapter 7

Concluding remarks

In this thesis, I have derived and tested a simple model for modelling the velocity of matter around cosmic filaments using linear theory and compared it to data from numerical simulations. This experiment is interesting as it can potentially allow for an additional set of experiments for extracting valuable cosmological information from the density field. I have also tested a proposed improvement to the radial velocity of matter around voids and applied it to a model of the void-galaxy cross-correlation function using a linear bias approximation. This was compared to the regular model for the radial velocity derived from linear theory. The analysis conducted on voids is useful as future observational datasets will grow in size. This will in turn require a better understanding of potential errors and improvements for current methodology. For these velocity models, the quadrupole of the void-galaxy cross-correlation was compared by applying cuts to the dataset including voids of different sizes in order to see if this affected the validity of the velocity approximations. Lastly, using a dark matter overdensity profile from another dataset, the regular velocity model was tested to see if potential corrections could be applied to an arbitrary density profile while still providing accurate parameter fits.

7.1 Summary of main results

Using linear theory, a simple model for the radial velocity perpendicular to the spine of cosmological filaments, similar to the model already tested on voids, was derived and tested. Overall this model looked promising, however, the comparison with numerical data illustrated the presence of irregularities regarding the amplitude of the velocity profile close to the filament spine. The amplitude of the stacked radial velocity profile for filaments showed large deviations depending on the dataset and σ -threshold chosen as input to the DisPerSE code. The promising aspect of the model is that it accurately replicated the shape of the numerically calculated velocity profile. The modelling of the radial velocity profile for filaments is reliant on accurate modelling of the density profile. The density profile for the filaments showed varying densities close to the filament spine depending on the dataset and σ -threshold chosen. This may be an effect

of how DisPerSE applies cuts to the filament catalogue removing filaments of different lengths, which in turn may affect the number density in the stacked density profile. It may also be a reasonable result, and in turn, will suggest that linear theory does not provide a reasonable approximation to the radial velocity of matter around filaments. However, on the basis of the results presented in this thesis, although the results looked promising, one is not able to draw a conclusion of whether or not linear theory is a sufficient approximation for the behavior of matter around filaments.

In this thesis, the regular linear model for the radial velocity of matter around filaments was also tested. In addition, a proposed improvement to the regular velocity model was tested and both models were compared with numerical data. Both of these velocity models were implemented into the anisotropic void-galaxy cross-correlation function and their performance was tested in terms of fitting cosmological parameters. When conducting the comparison between the two parameters, different cuts to the dataset was applied. The ϵ parameter, in which directly linked to the cosmological parameters, received consistently accurate fits using the regular velocity model. The model with the velocity correction term was biased in the sense that it consistently underestimated fits to ϵ . On the other hand, the model with the correction term made better fits to the β parameter when small voids were considered. However, when only large voids were considered, the regular model consistently provided the best fits for both of these parameters for the two largest datasets. The modelling of the velocity itself showed that when applying cuts only keeping the larger voids, the regular model qualitatively resembled the numerically calculated radial velocity in an accurate fashion. When applying cuts including only smaller voids there was not as prominent of an all round difference, but both models seemed to capture certain features better. The regular model performed better at capturing the velocity on large scales, while the amplitude predicted by the model with correction was more accurate with these cuts. These results suggest that when applying cuts to the dataset, excluding small voids which may be subject to non-linear behavior, the regular velocity model performed better.

Finally, the model for the anisotropic cross-correlation function between voids and galaxies was tested using a dark matter overdensity profile gathered from another dataset. Using this overdensity profile only the regular velocity model was implemented. For this test, the cuts only considering smaller voids was excluded. Across all datasets and cuts, this model managed to provide accurate samples for the ϵ parameter. Using no cuts, the model provided unsatisfying samples for $f\sigma_8$. This improved however when cuts only considering larger voids was considered, but the sampling was still not accurate enough for this implementation of the model to be deemed sufficient. This suggests that the model was not able to sample the amplitude of dark matter density profile accurately. The scaling of the overdensity profile, parametrized by the r_{scale} parameter, did however show that the model captured in which direction the density profile should be shifted. Overall, this analysis suggests that when sampling the ϵ parameter, using an arbitrary dark matter overdensity profile and correcting it may be sufficient. However, if one wants to sample $f\sigma_8$, the methodology applied in this analysis may not be sufficient.

7.2 Suggestions for future work

Although this thesis showed promising results when modelling the radial velocity of matter around filaments, further research is needed in order to conclude whether the approach proposed in this thesis may prove to be a full-fledged addition to LSS studies. The results showed that changing σ -threshold as input to the DisPerSE code effectively lowered the amount of shorter filaments. This also affected the amplitude of the radial velocity predicted by the model. In my analysis, I chose to apply cuts including filaments with length $20\text{Mpc}/h < l < 100\text{Mpc}/h$. This is a rather generous cut and applying more cuts, similar to what was done when studying voids, could indicate whether modelling the radial velocity using linear theory is a better approximation for filaments of different length. The model was also only tested on the smaller MD1 and MD2 datasets. Applying the model to the larger MD3 or MD4 datasets could also prove valuable. However, the method itself for calculating the numerical stacked velocity is computationally heavy when the number of particles increases. This is partly due to an increasing amount of cross products needed to gain the radial component of the velocity vector. If one is to judge whether the model sufficiently captures the behavior of the stacked radial velocity then this also has to be calculated. It could therefore be useful to implement the model in a more efficient manner than a pure Python implementation. The velocity profiles also showed a feature resembling a hook. As discussed, this feature may be attributed to most of the matter around filaments being contained at intersections where filaments meet. In order to test whether this is the case, a 2D modelling of the velocity field around filaments would be an interesting addition to the 1D radial velocity. A natural step once the modelling of the radial velocity has been tested, and in case it is deemed to be sufficient approximation, is to derive an expression for the filament-galaxy cross-correlation function in redshift space, similar to what already has been done for the void-galaxy cross-correlation function. This can potentially open up for future analysis similar to the what is conducted on voids in this thesis.

The results for the void galaxy cross-correlation function suggested that the preferable way to model the radial velocity of matter around voids was the regular model derived from linear theory. However, using a linear bias approximation, this model did not manage to accurately sample the β parameter when small voids were included in the catalogue. The sampling did however improve when cuts were applied. The $r > 40\text{Mpc}/h$ cut however could be viewed as rather conservative. Further effort could also be invested into applying cuts including smaller voids and try to better understand for what scales linear-theory is still a good approximation. This will increase the sample size of void-galaxy pairs and will in turn reduce noise in the numerically calculated cross-correlation function. The model with the velocity correction did however fit the β parameter better on smaller voids. Adding an extra dimensionless nuisance parameter as a scaling for the correction term could be a reasonable improvement to this particular model.

The model with the dark matter overdensity profile managed to provide satisfying fits the ϵ parameter across all datasets and cuts. It did not manage to sample $f\sigma_8$.

Applying cuts, however, did make the sampling of the parameter better. Applying the $r > 40\text{Mpc}/h$ cut did reduce the number of void-halo pairs and introduced noise to the cross-correlation function. A future suggestion could be to use even more conservative cuts. However, in order to do this, a larger dataset may be necessary in order to increase the samples of void-halo pairs. Performing the same analysis with the complete information about δ_{dm} corresponding to the given dataset could also prove valuable. A natural next step for all the methods tested, including the analysis of filaments, would be to use the redshifted data catalogue and apply the reconstruction algorithm and see how this affects the model results.

Bibliography

- [1] N. Aghanim, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, and et al. Planck 2018 results. *Astronomy & Astrophysics*, 641:A1, Sep 2020.
- [2] Volker Springel, Simon D. M. White, Adrian Jenkins, Carlos S. Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, John A. Peacock, Shaun Cole, Peter Thomas, Hugh Couchman, August Evrard, Jörg Colberg, and Frazer Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435(7042):629–636, June 2005.
- [3] Seshadri Nadathur, Will J. Percival, Florian Beutler, and Hans A. Winther. Testing low-redshift cosmic acceleration with large-scale structure. *Physical Review Letters*, 124(22), Jun 2020.
- [4] Ixandra Achitouv. Improved model of redshift-space distortions around voids: Application to quintessence dark energy. *Phys. Rev. D*, 96(8):083506, October 2017.
- [5] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421, July 1965.
- [6] George F. Smoot. Cobe observations and results. *Conference on 3K cosmology*, 1999.
- [7] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. *ApJS*, 208(2):19, October 2013.
- [8] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, E. Di Valentino, J. M. Diego,

- O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, Y. Fantaye, M. Farhang, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Kesitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, P. Lemos, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, L. Montier, G. Morgante, A. Moss, P. Natoli, H. U. Nørgaard-Nielsen, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Toffolatti, M. Tomasi, T. Trombetti, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca. Planck 2018 results. VI. Cosmological parameters. *A&A*, 641:A6, September 2020.
- [9] D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan, A. G. Riess, S. Rodney, E. Berger, D. J. Brout, P. J. Challis, M. Drout, D. Finkbeiner, R. Lunnan, R. P. Kirshner, N. E. Sanders, E. Schlafly, S. Smartt, C. W. Stubbs, J. Tonry, W. M. Wood-Vasey, M. Foley, J. Hand, E. Johnson, W. S. Burgett, K. C. Chambers, P. W. Draper, K. W. Hodapp, N. Kaiser, R. P. Kudritzki, E. A. Magnier, N. Metcalfe, F. Bresolin, E. Gall, R. Kotak, M. McCrum, and K. W. Smith. The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. *ApJ*, 859(2):101, June 2018.
- [10] J. Richard Bond, Lev Kofman, and Dmitry Pogosyan. How filaments of galaxies are woven into the cosmic web. *Nature*, 380(6575):603–606, 1996.
- [11] P. J. E. Peebles and J. T. Yu. Primeval adiabatic perturbation in an expanding universe. *Astrophys. J.*, 162:815–836, 1970.
- [12] P. J. E. Peebles. *The large-scale structure of the universe*. Princeton University Press, 1980.

- [13] Daniel J. Eisenstein, Idit Zehavi, David W. Hogg, Roman Scoccimarro, Michael R. Blanton, Robert C. Nichol, Ryan Scranton, Hee-Jong Seo, Max Tegmark, Zheng Zheng, and et al. Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies. *The Astrophysical Journal*, 633(2):560–574, Nov 2005.
- [14] Tröster, Tilman, Sánchez, Ariel. G., Asgari, Marika, Blake, Chris, Crocce, Martín, Heymans, Catherine, Hildebrandt, Hendrik, Joachimi, Benjamin, Joudaki, Shahab, Kannawadi, Arun, Lin, Chieh-An, and Wright, Angus. Cosmology from large-scale structure - constraining with boss. *A&A*, 633:L10, 2020.
- [15] Julian E Bautista, Romain Paviot, Mariana Vargas Magaña, Sylvain de la Torre, Sebastien Fromenteau, Hector Gil-Marín, Ashley J Ross, Etienne Burtin, Kyle S Dawson, Jiamin Hou, and et al. The completed sdss-iv extended baryon oscillation spectroscopic survey: measurement of the bao and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1. *Monthly Notices of the Royal Astronomical Society*, 500(1):736–762, Sep 2020.
- [16] Edwin Hubble. A relation between distance and radial velocity among extragalactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173, 1929.
- [17] Seshadri Nadathur and Will J Percival. An accurate linear model for redshift space distortions in the void–galaxy correlation function. *Monthly Notices of the Royal Astronomical Society*, 483(3):3472–3487, Dec 2018.
- [18] Seshadri Nadathur, Paul M. Carter, Will J. Percival, Hans A. Winther, and Julian E. Bautista. Beyond bao: Improving cosmological constraints from boss data with measurement of the void-galaxy cross-correlation. *Phys. Rev. D*, 100:023504, Jul 2019.
- [19] Noam I. Libeskind, Rien van de Weygaert, Marius Cautun, Bridget Falck, Elmo Tempel, Tom Abel, Mehmet Alpaslan, Miguel A. Aragón-Calvo, Jaime E. Forero-Romero, Roberto Gonzalez, and et al. Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473(1):1195–1217, Aug 2017.
- [20] Daniel J. Eisenstein, David H. Weinberg, Eric Agol, Hiroaki Aihara, Carlos Allende Prieto, Scott F. Anderson, James A. Arns, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, and et al. Sdss-iii: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems. *The Astronomical Journal*, 142(3):72, Aug 2011.
- [21] R. Laureijs, J. Amiaux, S. Arduini, J. L. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, B. Garilli, P. Gondoin, L. Guzzo, J. Hoar, H. Hoekstra, R. Holmes, T. Kitching, T. Maciaszek, Y. Mellier, F. Pasian,

- W. Percival, J. Rhodes, G. Saavedra Criado, M. Sauvage, R. Scaramella, L. Valenziano, S. Warren, R. Bender, F. Castander, A. Cimatti, O. Le Fèvre, H. Kurki-Suonio, M. Levi, P. Lilje, G. Meylan, R. Nichol, K. Pedersen, V. Popa, R. Rebolo Lopez, H. W. Rix, H. Rottgering, W. Zeilinger, F. Grupp, P. Hudelot, R. Massey, M. Meneghetti, L. Miller, S. Paltani, S. Paulin-Henriksson, S. Pires, C. Saxton, T. Schrabback, G. Seidel, J. Walsh, N. Aghanim, L. Amendola, J. Bartlett, C. Baccigalupi, J. P. Beaulieu, K. Benabed, J. G. Cuby, D. Elbaz, P. Fosalba, G. Gavazzi, A. Helmi, I. Hook, M. Irwin, J. P. Kneib, M. Kunz, F. Mannucci, L. Moscardini, C. Tao, R. Teyssier, J. Weller, G. Zamorani, M. R. Zapatero Osorio, O. Boulade, J. J. Foumond, A. Di Giorgio, P. Guttridge, A. James, M. Kemp, J. Martignac, A. Spencer, D. Walton, T. Blümchen, C. Bonoli, F. Bortoletto, C. Cerna, L. Corcione, C. Fabron, K. Jahnke, S. Ligori, F. Madrid, L. Martin, G. Morgante, T. Pamplona, E. Prieto, M. Riva, R. Toledo, M. Trifoglio, F. Zerbi, F. Abdalla, M. Douspis, C. Grenet, S. Borgani, R. Bouwens, F. Courbin, J. M. Delouis, P. Dubath, A. Fontana, M. Frailis, A. Grazian, J. Koppenhöfer, O. Mansutti, M. Melchior, M. Mignoli, J. Mohr, C. Neissner, K. Noddle, M. Poncet, M. Scodreggio, S. Serrano, N. Shane, J. L. Starck, C. Surace, A. Taylor, G. Verdoes-Kleijn, C. Vuerli, O. R. Williams, A. Zacchei, B. Altieri, I. Escudero Sanz, R. Kohley, T. Oosterbroek, P. Astier, D. Bacon, S. Bardelli, C. Baugh, F. Bellagamba, C. Benoist, D. Bianchi, A. Biviano, E. Branchini, C. Carbone, V. Cardone, D. Clements, S. Colombi, C. Conselice, G. Cresci, N. Deacon, J. Dunlop, C. Fedeli, F. Fontanot, P. Franzetti, C. Giocoli, J. Garcia-Bellido, J. Gow, A. Heavens, P. Hewett, C. Heymans, A. Holland, Z. Huang, O. Ilbert, B. Joachimi, E. Jennins, E. Kerins, A. Kiessling, D. Kirk, R. Kotak, O. Krause, O. Lahav, F. van Leeuwen, J. Lesgourgues, M. Lombardi, M. Magliocchetti, K. Maguire, E. Majerotto, R. Maoli, F. Marulli, S. Maurogordato, H. McCracken, R. McLure, A. Melchiorri, A. Merson, M. Moresco, M. Nonino, P. Norberg, J. Peacock, R. Pello, M. Penny, V. Pettorino, C. Di Porto, L. Pozzetti, C. Quercellini, M. Radovich, A. Rassat, N. Roche, S. Ronayette, E. Rossetti, B. Sartoris, P. Schneider, E. Semboloni, S. Serjeant, F. Simpson, C. Skordis, G. Smadja, S. Smartt, P. Spano, S. Spiro, M. Sullivan, A. Tilquin, R. Trotta, L. Verde, Y. Wang, G. Williger, G. Zhao, J. Zoubian, and E. Zucca. Euclid definition study report, 2011.
- [22] Anatoly Klypin, Gustavo Yepes, Stefan Gottlöber, Francisco Prada, and Steffen Heß. MultiDark simulations: the story of dark matter halo concentrations and density profiles. *MNRAS*, 457(4):4340–4359, April 2016.
- [23] P Schneider. *Extragalactic astronomy and cosmology : an introduction*. Springer, Springer-Verlag Berlin Heidelberg, 2015.
- [24] Scott Dodelson. *Modern cosmology*. Academic Press, San Diego, CA, 2003.
- [25] Barbara Ryden. *Introduction to cosmology*. Cambridge University Press, New York, NY, 2017.

- [26] Gaia Collaboration A.G.A. Brown, A. Vallenari, T. Prusti, J. Bruijne, C. Babusi-
aux, M. B. L. Observatory, L. University, Inaf Padova, ESA-Estec, Gepi, Obser-
vatoire de Paris, Univ. Grenoble Alpes, A. Recheninstitut, and Heidelberg. Gaia
early data release 3: Summary of the contents and survey properties. 2020.
- [27] Jaswant Yadav, Somnath Bharadwaj, Biswajit Pandey, and T. R. Seshadri. Testing
homogeneity on large scales in the Sloan Digital Sky Survey Data Release One.
Monthly Notices of the Royal Astronomical Society, 364(2):601–606, 12 2005.
- [28] A Einstein. Die feldgleichungen der gravitation. *Sitzungsberichte der Königlich
Preußischen Akademie der Wissenschaften (Berlin)*, pages 844–847, 1915.
- [29] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini,
A. J. Banday, R. B. Barreiro, N. Bartolo, and et al. Planck 2018 results. *Astronomy
& Astrophysics*, 641:A6, Sep 2020.
- [30] A. Einstein and W. de Sitter. On the Relation between the Expansion and the
Mean Density of the Universe. *Proceedings of the National Academy of Science*,
18(3):213–214, March 1932.
- [31] Gerson Goldhaber and David B. Cline. The acceleration of the expansion of the
universe: A brief early history of the supernova cosmology project (scp). *AIP
Conference Proceedings*, 2009.
- [32] Alexei V. Filippenko and Adam G. Riess. Results from the high-z supernova search
team. *Physics Reports*, 307(1-4):31–44, Dec 1998.
- [33] P J E Peebles. Structure of the coma cluster of galaxies. *The Astronomical journal.*,
75:13, 1970.
- [34] Giovanni F Fazio, Rennan Barkana, Shinji Tsujikawa, Jihn E Kim, and Kentaro
Nagamine. *The encyclopedia of cosmology*. World Scientific, Singapore, 2018.
- [35] Mark Holmes. *Introduction to numerical methods in differential equations*. Springer,
New York, 2007.
- [36] R. W Hockney. Computer simulation using particles, 1981.
- [37] Piet Hut and Josh Barnes. A hierarchical $O(n \log n)$ force-calculation algorithm.
Nature (London), 324(6096):446–449, 1986.
- [38] Andy Lawrence. *Probability in Physics, An Introductory Guide*. Springer Nature
Switzerland, 2019.
- [39] Vincent.J. Martinez, Enn. Saar, Enrique Martinèz. Gonzales, and Maria-Jesús
Pons-Boderia. *Data Analysis in Cosmology*. Springer-Verlag Berlin, 2009.
- [40] Alan Heavens. Statistical techniques in cosmology, 2010.

- [41] T. Sousbie. The persistent cosmic web and its filamentary structure - I. Theory and implementation. *MNRAS*, 414:350–383, June 2011.
- [42] T. Sousbie, C. Pichon, and H. Kawahara. The persistent cosmic web and its filamentary structure - II. Illustrations. *MNRAS*, 414:384–403, June 2011.
- [43] Alex Ho. Probing Modified Gravity in Cosmic Filaments . 2018.
- [44] J. Milnor, M. SPIVAK, and R. WELLS. *Morse Theory. (AM-51), Volume 51*. Princeton University Press, 1969.
- [45] Robin Forman. Morse theory for cell complexes. *Advances in Mathematics*, 134(1):90 – 145, 1998.
- [46] W. E. Schaap and R. van de Weygaert. Continuous fields and discrete samples: reconstruction through Delaunay tessellations. *A&A*, 363:L29–L32, November 2000.
- [47] Edelsbrunner, Letscher, and Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, 2002.
- [48] Mark C. Neyrinck. zobov: a parameter-free void-finding algorithm. *Monthly Notices of the Royal Astronomical Society*, 386(4):2101–2109, Jun 2008.
- [49] Seshadri Nadathur, Paul Carter, and Will J Percival. A zeldovich reconstruction method for measuring redshift space distortions using cosmic voids. *Monthly Notices of the Royal Astronomical Society*, 482(2):2459–2470, Oct 2018.
- [50] A. Burden, W. J. Percival, and C. Howlett. Reconstruction in Fourier space. *Monthly Notices of the Royal Astronomical Society*, 453(1):456–468, 08 2015.
- [51] Adi Nusser and Marc Davis. On the Prediction of Velocity Fields from Redshift Space Galaxy Samples. *ApJ*, 421:L1, January 1994.
- [52] Lars Hernquist and Neal Katz. TREESPH: A Unification of SPH with the Hierarchical Tree Method. *ApJS*, 70:419, June 1989.
- [53] Volker Springel. The cosmological simulation code GADGET-2. *MNRAS*, 364(4):1105–1134, December 2005.
- [54] Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores. *The Astrophysical Journal*, 762(2):109, Dec 2012.
- [55] Paolo Catelan, Francesco Lucchin, Sabino Matarrese, and Lauro Moscardini. Eulerian perturbation theory in non-flat universes: second-order approximation. *Monthly Notices of the Royal Astronomical Society*, 276(1):39–56, 09 1995.
- [56] David Alonso. Cute solutions for two-point correlation functions from large cosmological datasets, 2013.

- [57] Stephen D. Landy and Alexander S. Szalay. Bias and Variance of Angular Correlation Functions. *ApJ*, 412:64, July 1993.
- [58] Nico Hamaus, Marie-Claude Cousinou, Alice Pisani, Marie Aubert, Stéphanie Escoffier, and Jochen Weller. Multipole analysis of redshift-space distortions around cosmic voids. *Journal of Cosmology and Astroparticle Physics*, 2017(07):014–014, Jul 2017.
- [59] C. Alcock and B. Paczynski. An evolution free test for non-zero cosmological constant. *Nature*, 281:358, October 1979.
- [60] Kyle S. Dawson, David J. Schlegel, Christopher P. Ahn, Scott F. Anderson, Éric Aubourg, Stephen Bailey, Robert H. Barkhouser, Julian E. Bautista, Alessandra Beifiori, Andreas A. Berlind, and et al. The baryon oscillation spectroscopic survey of sdss-iii. *The Astronomical Journal*, 145(1):10, Dec 2012.
- [61] Francisco-Shu Kitaura, Sergio Rodríguez-Torres, Chia-Hsun Chuang, Cheng Zhao, Francisco Prada, Héctor Gil-Marín, Hong Guo, Gustavo Yepes, Anatoly Klypin, Claudia G. Scóccola, Jeremy Tinker, Cameron McBride, Beth Reid, Ariel G. Sánchez, Salvador Salazar-Albornoz, Jan Niklas Grieb, Mariana Vargas-Magana, Antonio J. Cuesta, Mark Neyrinck, Florian Beutler, Johan Comparat, Will J. Percival, and Ashley Ross. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release. *MNRAS*, 456(4):4156–4173, March 2016.
- [62] Sergio A. Rodríguez-Torres, Chia-Hsun Chuang, Francisco Prada, Hong Guo, Anatoly Klypin, Peter Behroozi, Chang Hoon Hahn, Johan Comparat, Gustavo Yepes, Antonio D. Montero-Dorta, Joel R. Brownstein, Claudia Maraston, Cameron K. McBride, Jeremy Tinker, Stefan Gottlöber, Ginevra Favole, Yiping Shu, Francisco-Shu Kitaura, Adam Bolton, Román Scoccimarro, Lado Samushia, David Schlegel, Donald P. Schneider, and Daniel Thomas. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: modelling the clustering and halo occupation distribution of BOSS CMASS galaxies in the Final Data Release. *MNRAS*, 460(2):1173–1187, August 2016.
- [63] I. Horvath, Z. Bagoly, J. Hakkila, and L. V. Toth. Anomalies in the GRBs’ spatial distribution. In *Proceedings of Swift: 10 Years of Discovery (SWIFT 10)*, page 78, December 2014.
- [64] Daniela Galárraga-Espinosa, Nabila Aghanim, Mathieu Langer, Céline Gouin, and Nicola Malavasi. Populations of filaments from the distribution of galaxies in numerical simulations. *Astronomy & Astrophysics*, 641:A173, Sep 2020.
- [65] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The illustriatng simulations: Public data release, 2021.

-
- [66] Bonjean, V., Aghanim, N., Douspis, M., Malavasi, N., and Tanimura, H. Filament profiles from wisexscos galaxies as probes of the impact of environmental effects. *A&A*, 638:A75, 2020.
- [67] Maciej Bilicki, John A. Peacock, Thomas H. Jarrett, Michelle E. Cluver, Natasha Maddox, Michael J. I. Brown, Edward N. Taylor, Nigel C. Hambly, Aleksandra Solarz, Benne W. Holwerda, Ivan Baldry, Jon Loveday, Amanda Moffett, Andrew M. Hopkins, Simon P. Driver, Mehmet Alpaslan, and Joss Bland-Hawthorn. WISEX SuperCOSMOS PHOTOMETRIC REDSHIFT CATALOG: 20 MILLION GALAXIES OVER 3π STERADIANS. *The Astrophysical Journal Supplement Series*, 225(1):5, jul 2016.